

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 16, 2005, 21:55:52 ; Search time 210.545 Seconds  
(without alignments) 17070.053 Million cell updates/sec

Title: US-09-551-494-5\_COPY\_5430\_5505  
 Perfect score: 76  
 Sequence: 1 qtgcagacggctcgccaat.....tgaagtaccaatggctgtga 76

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

```
Database : GenEmbl.*
1: gb_bai.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES						
Result No.	Score	Query Match	Length	DB	ID	Description
1	76	100.0	6355	6	AX040177	Sequence
2	76	100.0	6355	6	AX795380	AX795380 Sequence
3	76	100.0	6355	14	TMGGC	M34077 Tobacco mil
4	71.2	93.7	1413	14	TMGMPCPA	M34236 Tobacco mil
5	69.6	91.6	6356	14	AB078435	M3478435 Tobacco m
6	58.6	77.1	411	6	AX040184	Sequence
7	45.8	60.3	912	14	ORVTRANS	X55296 Odontogloss
8	45.8	60.3	5997	6	E03624	E03624 DNA encodin
9	45.8	60.3	6597	6	E04305	E04305 cDNA encodi
10	45.8	60.3	6609	14	ORU34586	U34586 Odontogloss
11	45.8	60.3	6611	14	S83257	S83257 126 kda pro
12	45.8	60.3	6612	14	AY571290	AY571290 Odontoglo
13	44	57.9	1004	14	TMVW2	V01406 TMV 3' end.
14	44	57.9	1019	6	E00088	E00088 Genomic RNA
15	44	57.9	6395	14	D63809	D63809 Tobacco mos
16	42.4	55.8	396	14	TMQ307579	AJ307579 Tobacco m
17	42.4	55.8	632	6	A68760	A68760 Sequence 1
18	42.4	55.8	706	14	TM308689	AJ308689 Tobacco m
19	42.4	55.8	750	14	TMQ509080	AJ509080 Tobacco m

20	42.4	55.8	750	14	TWO509084	AJ509084	Tobacco m
21	42.4	55.8	782	14	TWO308692	AJ308692	Tobacco m
22	42.4	55.8	788	14	TWO308685	AJ308685	Tobacco m
23	42.4	55.8	790	14	TWO307583	AJ307583	Tobacco m
24	42.4	55.8	804	6	CQ793025	Sequence	
25	42.4	55.8	804	14	TWO307582	AJ307582	Tobacco m
26	42.4	55.8	806	14	TWO308682	AJ308682	Tobacco m
27	42.4	55.8	806	14	TWO308684	AJ308684	Tobacco m
28	42.4	55.8	807	6	BD263879	BD263879	Viral exp
29	42.4	55.8	807	6	BD263880	BD263880	Viral exp
30	42.4	55.8	807	6	AR435699	AR435699	Sequence
31	42.4	55.8	807	6	AR435700	AR435700	Sequence
32	42.4	55.8	807	6	AX045758	AX045758	Sequence
33	42.4	55.8	807	6	AX592974	AX592974	Sequence
34	42.4	55.8	807	14	TWO307581	AJ307581	Tobacco m
35	42.4	55.8	807	14	TWO308683	AJ308683	Tobacco m
36	42.4	55.8	807	14	TWO308693	AJ308693	Tobacco m
37	42.4	55.8	807	14	TWO509081	AJ509081	Tobacco m
38	42.4	55.8	808	14	TWO307578	AJ307578	Tobacco m
39	42.4	55.8	808	14	TWO308686	AJ308686	Tobacco m
40	42.4	55.8	808	14	TWO308688	AJ308688	Tobacco m
41	42.4	55.8	808	14	TWO308690	AJ308690	Tobacco m
42	42.4	55.8	809	14	TWO308691	AJ308691	Tobacco m
43	42.4	55.8	811	14	TWO509082	AJ509082	Tobacco m
44	42.4	55.8	815	14	AX300161	AY300161	Tobacco m
45	42.4	55.8	817	14	TWO509083	AJ509083	Tobacco m

ALIGNMENTS

RESULT 1

AX040177

LOCUS

Sequence 5 from Patent WO0063397.

AX040177

VERSION

AX040177.1

GI:11230127

KEYWORDS

synthetic construct

SOURCE

artificial sequences.

ORGANISM

1

REFERENCE

AUTHORS

Meulewaeter, F., Cornelisse, Jacobs, J., van Eldik, G. and Metzlaaff, M.

TITLE

Methods and means for delivering inhibitory rna to plants and applications thereof

JOURNAL

Patent: WO 0063397-A 5 26-OCT-2000,

FEATURES

source

Location/Qualifiers

1. .6355

/organism="synthetic construct"

/mol\_type="unassigned DNA"

/db\_xref="taxon:32630"

/note="cDNA copy of the nucleotide sequence of the genome of TMV-U2"

ORIGIN

Query Match 100.0%; Score 76; DB 6; Length 6355;

Best Local Similarity 100.0%; Pred. No. 2.2e-15;

Matches 76; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTGACAGCGCTCGCCAATTGAACCTACTGAAAGCTTTGTGAGAGGTTCTGGGATGAA 60

Db 5430 GTGACAGCGCTCGCCAATTGAACCTACTGAAAGCTTTGTGAGAGGTTCTGGGATGAA 5489

Qy 61 GTACCAATGGCTGTGA 76

Db 5490 GTACCAATGGCTGTGA 5505

RESULT 2

AX795380

LOCUS

Sequence 7 from Patent WO03052108.

AX795380

VERSION

AX795380

KEYWORDS

6355 bp

SOURCE

mRNA

ORGANISM

1

REFERENCE

AUTHORS

Meulewaeter, F., Cornelisse, Jacobs, J., van Eldik, G. and Metzlaaff, M.

TITLE

Methods and means for delivering inhibitory rna to plants and applications thereof

JOURNAL

Patent: WO 0063397-A 5 26-OCT-2000,

FEATURES

source

Location/Qualifiers

1. .6355

/organism="synthetic construct"

/mol\_type="unassigned DNA"

/db\_xref="taxon:32630"

/note="cDNA copy of the nucleotide sequence of the genome of TMV-U2"

ORIGIN

Query Match 100.0%; Score 76; DB 6; Length 6355;

Best Local Similarity 100.0%; Pred. No. 2.2e-15;

Matches 76; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTGACAGCGCTCGCCAATTGAACCTACTGAAAGCTTTGTGAGAGGTTCTGGGATGAA 60

Db 5430 GTGACAGCGCTCGCCAATTGAACCTACTGAAAGCTTTGTGAGAGGTTCTGGGATGAA 5489

Qy 61 GTACCAATGGCTGTGA 76

Db 5490 GTACCAATGGCTGTGA 5505

```
ACCESSION AX795380
VERSION AX795380.1 GI:37516053
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM artificial sequences.
REFERENCE 1
AUTHORS Metzlauff, M.H., Gossele, V.M., Meulewaeter, F. and Fache, I.C.
TITLE Improved methods and means for delivering inhibitory rna to plants
JOURNAL Patent: WO 03052108-A 7 26-JUN-2003;
Bayer Bioscience N.V. (BE)
FEATURES
    source
        1..6355
            /organism="synthetic construct"
            /mol_type="mRNA"
            /db_xref="taxon:32630"
            /note="cDNA sequence of the genome of TMV-U2"
ORIGIN
    Query Match 100.0%; Score 76; DB 6; Length 6355;
    Best Local Similarity 100.0%; Pred. No. 2.2e-15;
    Matches 76; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
    QY 1 GTGACAGCGGTCGCGCAATTCGAATCACTCACTGAAAGGTTGTTGAGGAGTTCGTGGATGAA 60
    Db 5430 GTGACAGCGGTCGCGCAATTCGAATCACTCACTGAAAGGTTGTTGAGGAGTTCGTGGATGAA 5489
    QY 61 GTACCAATGGCTGTGA 76
    Db 5490 GTACCAATGGCTGTGA 5505
RESULT 3
TMCGG
LOCUS Tobacco mild green mosaic virus complete genome. 6355 bp ss-RNA linear VRL 03-AUG-1993
DEFINITION Tobacco mild green mosaic virus complete genome.
ACCESSION M34077.1 M22483
VERSION M34077.1 GI:335243
KEYWORDS coat protein.
SOURCE Tobacco mild green mosaic virus
ORGANISM Tobacco mild green mosaic virus
REFERENCE 1 (bases 6127 to 6355)
AUTHORS Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
TITLE Garcia-Arenal, F.
SEQUENCE AND STRUCTURE AT THE GENOME 3' END OF THE U2-STRAIN OF
Tobacco mosaic virus, a histidine-accepting tobamovirus
JOURNAL Virology 167 (1), 201-206 (1998)
MEDLINE 89045644
PUBMED 3188396
REFERENCE 2 (bases 1 to 6355)
AUTHORS Solis, I. and Garcia-Arenal, F.
TITLE The complete nucleotide sequence of the genomic RNA of the
tobamovirus tobacco mild green mosaic virus
JOURNAL Virology 177 (2), 553-558 (1990)
MEDLINE 90320127
PUBMED 2371769
COMMENT Original source text: Tobacco mild green mosaic virus (strain
U2-TMV), cDNA to viral RNA, from N.tabacum cv. Samsum.
Draft entry and computer-readable sequence for [1] kindly submitted
by F. Garcia-Arenal, 10-FEB-1989. The RNA appears to have a RNA-
like, L-shaped structure at the 3' terminus, linked to a quasi-
continuous double-helical stalk, with five pseudoknots involved in
the formation of the whole structure. However, the structure of
U2-TMV RNA is less stringently conserved than the 3' termini of
'vulgar' and other histidine-accepting tobamoviruses. Draft entry
and computer-readable sequence for [1] kindly submitted by
F. Garcia-Arenal, 08-MAY-1990, for release after publication.
Location/Qualifiers
    1..6355
        /organisms="Tobacco mild green mosaic virus"
        /mol_type="genomic RNA"
        /db_xref="taxon:12241"
```

CDS

```
join(71..3403,3407..4900)
/note="183 kDa protein"
/codon_start=1
/protein_id="AAA47934.1"
/db_xref="GI:335244"
/translation="MAHIQSIISNALLESVSGKNTLVNDLARRRMVDTAVEEFNARDR
SPKVNFSKTISEBOTLLVSNAYPEFOITFYNTQNAVHSLAGGLALELYMLQVPG
RPTVDIGNFAAHLFKRGDYVHCMPNLDIRDNRHGGQKDSIMYLSRLSRNKNVIP
EFQREAPNRYAEAPNEVCCKTFQDCRIHPENSGRRYAVALHSYLDIPVHBFGAALI
SKNHWCTAASILAELALDQTEVLNEIGATFKREGDDVFFFADESTLNYSKYN
ILHYVKSYPASSRIYVFKFELVTRNTWFKCTKVDYTYLYKSVQVGCSDQDFE
AMEDAFAYKKTLMFENTERAI FRDTASVNFWMFKMDWIVPLPFGSI TSKMTRFSE
IVNRDFVTVLHIRTQAKALTQNVLSFVESIRSRVINGVTARSEDWDVKAILQ
LSMTFFLOTKLAALODDI VMGKRCCLDKTTSELINDEVGKFGNVPFTIKERLVSRI
LDVSENALKIKIPDLVYTWKDRPVAEYTKSEELPHLDIKDLSEAEQMDALSLSIL
KGADNFDIAFKDMCKALDSDPVAARVIVAAENRSGLTITFDKPTENVAKALST
ASEAVVCLPTSEVTVNKFSAIEAKGRLPVCAESHGLTNANLEHQLSESLNDFHAKCV
DSVITTKWLLKPAKGHSGVVDYKGMFTALLSYEGDRMVTESDMDRVAVSSDT
GVMVDIAKQNLAKRTMRDGPHEPTAKMVLVDGVPKGYKGFERDLDEDLILVPG
MYSDIARLRANS SGLIRATMDNVTVDSLMLHPKPRSHKRLFIDEGMLHMGCVNFL
KQAAMTIRANSSGGLIRATMDNVTVDSLMLHPKPRSHKRLFIDEGMLHMGCVNFL
GLLENLIVAMIKRNFAPDLTGTDIESTASVVVDKFDVYIKKEKYTKNIAGVMTKD
SMRWLENKEVLLDLNANFTDLPAIDQYKMIKAQPKQKLDLSIQNSYPALQYK
YHSKQINGILAGSELTLLLEAFDSKFLFTRKTPTEQIQEFPDLDHVPMDVLEL
DISYDKSQNEFHCAVEYIWKRLGNEFLAEVWKQHRKTLTKDIAGIKCTKLTQYR
KSGDVTTFGNTVIIAACLGSMLPMKVIKGAFCGDDSVLYFPKGLDFDPIOCANLM
WNFEAKLYKRYGYGCRVYIHHDKGAI VYDPLKLSKLGAKHI KDYDHLERLYSL
CDVACSLGNWGLFQPQLNAAIKVHKTADIGSFAFNCVANKFLCKFLFRTPLNGC"
```

CDS

```
71..3406
/note="126 kDa protein"
/codon_start=1
/protein_id="AAA47935.1"
/db_xref="GI:335245"
/translation="MAHIQSIISNALLESVSGKNTLVNDLARRRMVDTAVEEFNARDR
SPKVNFSKTISEBOTLLVSNAYPEFOITFYNTQNAVHSLAGGLALELYMLQVPG
RPTVDIGNFAAHLFKRGDYVHCMPNLDIRDNRHGGQKDSIMYLSRLSRNKNVIP
EFQREAPNRYAEAPNEVCCKTFQDCRIHPENSGRRYAVALHSYLDIPVHBFGAALI
SKNHWCTAASILAELALDQTEVLNEIGATFKREGDDVFFFADESTLNYSKYN
ILHYVKSYPASSRIYVFKFELVTRNTWFKCTKVDYTYLYKSVQVGCSDQDFE
AMEDAFAYKKTLMFENTERAI FRDTASVNFWMFKMDWIVPLPFGSI TSKMTRFSE
IVNRDFVTVLHIRTQAKALTQNVLSFVESIRSRVINGVTARSEDWDVKAILQ
LSMTFFLOTKLAALODDI VMGKRCCLDKTTSELINDEVGKFGNVPFTIKERLVSRI
LDVSENALKIKIPDLVYTWKDRPVAEYTKSEELPHLDIKDLSEAEQMDALSLSIL
KGADNFDIAFKDMCKALDSDPVAARVIVAAENRSGLTITFDKPTENVAKALST
ASEAVVCLPTSEVTVNKFSAIEAKGRLPVCAESHGLTNANLEHQLSESLNDFHAKCV
DSVITTKWLLKPAKGHSGVVDYKGMFTALLSYEGDRMVTESDMDRVAVSSDT
GVMVDIAKQNLAKRTMRDGPHEPTAKMVLVDGVPKGYKGFERDLDEDLILVPG
MYSDIARLRANS SGLIRATMDNVTVDSLMLHPKPRSHKRLFIDEGMLHMGCVNFL
KQAAMTIRANSSGGLIRATMDNVTVDSLMLHPKPRSHKRLFIDEGMLHMGCVNFL
GLLENLIVAMIKRNFAPDLTGTDIESTASVVVDKFDVYIKKEKYTKNIAGVMTKD
SMRWLENKEVLLDLNANFTDLPAIDQYKMIKAQPKQKLDLSIQNSYPALQYK
YHSKQINGILAGSELTLLLEAFDSKFLFTRKTPTEQIQEFPDLDHVPMDVLEL
DISYDKSQNEFHCAVEYIWKRLGNEFLAEVWKQHRKTLTKDIAGIKCTKLTQYR
KSGDVTTFGNTVIIAACLGSMLPMKVIKGAFCGDDSVLYFPKGLDFDPIOCANLM
WNFEAKLYKRYGYGCRVYIHHDKGAI VYDPLKLSKLGAKHI KDYDHLERLYSL
CDVACSLGNWGLFQPQLNAAIKVHKTADIGSFAFNCVANKFLCKFLFRTPLNGC"
```

exon

```
<71..3403
/note="183 kDa protein"
```

exon

```
3407..4900
/note="183 kDa protein"
```

CDS

```
4890..5660
/note="28.5 kDa protein"
/codon_start=1
/protein_id="AAA47936.1"
/db_xref="GI:335246"
```

FEATURES

```
/translation="MAVSRDVTKISBFDLSKQDEILPAPMTKVSVRISTVVKIMA
VNDLSLSDVLLKGVKLVKGVKVLADLVSGEWNLPDNCGRGVSVCITVDRMRKRSKE
ATLGAYHAPACKNFSPKLI PNYSITSEDAEKHPQWLNI KGVAMEGYCPLSLEFV
SICVYHKNVNRKGLRRLISVTDGSPBELTEKVVEEFVDEVPMAVKLEBPENKKEMY
GNVNNKKNNSGKKGFKIEIEDVSDDESISSTP"
```

5666..6145

CDS

```

/notes="coat protein"
/codon_start=1
/protein_id="AAA47937.1"
/db_xref="GI:335247"
/translations="MPYITNSPSQFVYLSSAYADPVQLINLCTNALGNQFQTOQARTT
VQOQFADAWKFPVSMVTRPPASDFYVRYNSTLDPLITALLNSFDTRNRIIEVDNQA
PNTTEIVNATQRVDDATVAIRASINNLANELVRGTGMFNQAGFETASGLVWTTTPAT"

ORIGIN
Query Match      100.0%; Score 76; DB 14; Length 6355;
Best Local Similarity 100.0%; Pred. No. 2.2e-15;
Matches 76; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTGACAGACGGCTCGCAATTGAACCTCACTGAAAGGTTCTTGTGAGGAGTTCTGCGATGAA 60
    |||||
Db 5430 GTGACAGACGGCTCGCAATTGAACCTCACTGAAAGGTTCTTGTGAGGAGTTCTGCGATGAA 60
    |||||

Qy 61 GTACCAATGCGTGTGA 76
    |||||
Db 5490 GTACCAATGCGTGTGA 5505
    |||||

RESULT 4
TMGMPCPA 1413 bp ss-RNA linear VRL 03-AUG-1993
LOCUS Tobacco mild green mosaic virus movement and coat protein genes,
complete cds.
ACCESSION M34236
VERSION M34236.1 GI:335248
KEYWORDS coat protein; movement protein.
SOURCE Tobacco mild green mosaic virus
ORGANISM Tobacco mild green mosaic virus
REFERENCE 1 (bases 1 to 1413)
AUTHORS Nejidat,A., Cellier,F., Holt,C.A., Gafny,R., Eggenberger,A.L. and
Beachy,R.N.
TITLE Transfer of the movement protein gene between two tobamoviruses:
influence on local lesion development
JOURNAL Virology 180 (1), 318-326 (1991)
MEDLINE 91082424
PUBMED 1984654
COMMENT Original source text: Tobacco mild green mosaic virus (strain PV
228), cDNA to viral RNA.
Draft entry and computer-readable sequence for [1] kindly submitted
by A.Nejidat, 11-MAY-1990.
Location/Qualifiers
FEATURES
source
1. .1413
/organism="Tobacco mild green mosaic virus"
/mol_type="genomic RNA"
/db_xref="taxon:12241"
51. .851
/codon_start=1
/product="movement protein"
/protein_id="AAA47938.1"
/db_xref="GI:335249"
/translation="MAVSLRDTVKISFINLSKQDELLPAPMTKVKSVRISTYDKIMA
VKNDSDVDLLKGLVKNYGVLCGLVSVGWNLDPNCRGGVSVCIVDKRWKRNE
ATLGAYHAPACKKNFSLKIPNYSITSEDAEKNPQVLVNIKGVAMEEGCPLSLEFV
SICVWKNMVKGLRIELRVTDSDPIELTEKVVVEFVDPVPMVKLERPRKTKGKKK
RKEKKKRVVGVNSVNNKKINNKGKGLKVEIEDNVSDDESIASSSTF"
857. .1336
/codon_start=1
/product="coat protein"
/protein_id="AAA47939.1"
/db_xref="GI:335250"
/translation="MPYITNSPSQFVYLSSAYADPVQLINLCTNALGNQFQTOQARTT
VQOQFADAWKFPVSMVTRPPASDFYVRYNSTLDPLITALLNSFDTRNRIIEVDNQA
PNTTEIVNATQRVDDATVAIRASINNLANELVRGTGMFNQAGFETASGLVWTTTPAT"

CDS
Query Match      93.7%; Score 71.2; DB 14; Length 1413;
Best Local Similarity 96.1%; Pred. No. 7.6e-14;
Matches 73; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

ORIGIN
/notes="coat protein"
/codon_start=1
/protein_id="AAA47937.1"
/db_xref="GI:335247"
/translations="MPYITNSPSQFVYLSSAYADPVQLINLCTNALGNQFQTOQARTT
VQOQFADAWKFPVSMVTRPPASDFYVRYNSTLDPLITALLNSFDTRNRIIEVDNQA
PNTTEIVNATQRVDDATVAIRASINNLANELVRGTGMFNQAGFETASGLVWTTTPAT"

Qy 1 GTGACAGACGGCTCGCAATTGAACCTCACTGAAAGGTTCTTGTGAGGAGTTCTGCGATGAA 60
    |||||
Db 591 GTAACAGACGACTCGCAATTGAACCTCACTGAAAGGTTCTTGTGAGGAGTTCTGCGATGAA 650
    |||||

Qy 61 GTACCAATGCGTGTGA 76
    |||||
Db 651 GTACCAATGCGTGTGA 666
    |||||

RESULT 5
AB078435 6356 bp RNA linear VRL 22-JAN-2002
LOCUS Tobacco mild green mosaic virus complete genome, strain:Japanese.
DEFINITION
ACCESSION AB078435
VERSION AB078435.1 GI:18253266
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1
AUTHORS Okuno,T., Hamada,H., Takeuchi,S., Morishima,N., Yoshimoto,E. and
Hikichi,Y.
TITLE Nucleotide sequence of the Japanese isolate of Tobacco mild green
mosaic virus
JOURNAL
REFERENCE 2 (bases 1 to 6356)
AUTHORS Okuno,T., Hamada,H., Takeuchi,S., Morishima,N., Yoshimoto,E. and
Hikichi,Y.
TITLE Direct Submission
JOURNAL
REFERENCE 18-JAN-2002 Tetsuro Okuno, Kyoto University;
Kitsaishirakawa Oiwakecho, Sakyou 606-8502, Japan
(B-mail:okuno@kais.kyoto-u.ac.jp, Tel:81-75-753-6131,
Fax:81-75-753-6131)
Location/Qualifiers
FEATURES
source
1. .6356
/organism="Tobacco mild green mosaic virus"
/virion
/mol_type="genomic RNA"
/strain="Japanese"
/db_xref="taxon:12241"
/notes="RNA transcripts synthesized in vitro from the cDNA
clone of TMGMV-J direct systemic infection in Nicotiana
benthiana-TMGV-Japanese strain"
72. .4901
/notes="position 3405. 3407 is readthrough product"
/codon_start=1
/transl_except=1
/product="183kDa protein"
/protein_id="BAB83986.1"
/db_xref="GI:18253267"
/translation="MAHIQSTISNALLESVSGKNTLVNDLARRMYDTAVDEFNARDR
RPKNFPTISEEQTLVNTNAYPEFQITFYNTQNAVHSLAGLRALSELYMLQVPYG
SPTYDIGNFAHLFKGRDVKHCCMPNLDIRDIMRHEGQKDSIEMYLRLSRNKNVIP
EFORAAFNRYAABNPVCCSTQDQCIRIPEPNSGRYAIISLSDYIPVHEFGAALI
SKDIHCYAFSESEALLADQETVLTNETIGATKREGDVSPFADESTLYNHYKN
ILHYVKTSTPPASRIVYFEFLVTRVNTGATKREGDVSPFADESTLYNHYKN
AMEDATYKTLAMFNTERRAILDTSVNFPMKDMVLPVLFEGISTKMTRESE
IYNRDFVTLNHIYTHQKALTYQNVLSFVSIERSVINGVTARSEWDVDRKALQP
LSMTFLOTKLAALQDDIWMKFCRLDKTTSILIDWVGKFFGVNPTIKERLVSRIK
LDVSENAIKIKI PDLVYTWKDFVAEYKSEELPHLDIKKLEAEQMDALSLSIL
RDADNPDIKFKDMCKALCVNPDVAARVIVAVENBSGLTLTDPKTEENVAKAKST
ASEAVRLRPTSEVWVNFPSIAERGLPVCASSHGLTNLHSELESLSNDFHKACV
DSVITKOMASVVYTGSLKVKQMKYVDSLAASISATVSNLCKSLKQVGVSDSREKV
GWNVDLTKKLLKPAAGHSVGVLDYKGMFTALLSYEGDWMVTESDMRVAVSSDT
MYSDTAKLQNLKTRWDRGHEPTAKMVLVDGPGCGTKETILERVLDLDELTLVPG
KQAAITREANSSGLIRATMDNVRTVDSFLMHPKPRTHKRLTIDGLMLTGCVNFI
VLISGDIAYIGDTQIQIPEINRVQPPFPKHEKLOQVDEVMRRITLRCFQDVNFI
QSKYEGAVSTTSVQRSVSEMIKGGKGLNSVSKPLGKIVITQADKFELEKGYKN
VNTVHQETBEDVSLRLTATPLTILSKSPHVLVALTRTKTKFYKTVLDPVLQV
IISDLSLSFLLEMTWVAGKXQLQMDAVFKGHNLFVATPKSGDFPDQLQFYVDVCL
PGNSTITLKYDAVTMLRLNLSNLCVLDQDFSKIIPMPKEVKPCLEPLVATAEPV
AGLEENLVAMIKENFNAPDLTGTIDIESVAVVVDKFFDSYFLKKEKYTKNIAGVMTK
```

```
DSMRVLGKQERSTICQLANYNFDLPAIDQYKHKMIKAQPKQKLDLSTQNEYPALQTI
VHSKQINGFGFVSELRLLELAEVDSQKFLFTRKTEQIQEFSDIDSHVPMDFW
ELDISKIDQNEFHCAVEYEWKRLGLEAEVWVKQGRKTKLKYDAGTKCTLMWY
ORSGDVTFITGKVTIAACGLSMPLMEKVIKGAFCGDDSVLYFPGKLDFFDIOSCAN
LWNFPAKLVKGYGVFCGRYLIHDKGALVYDPLKLSKLGAKHIDYDHLSELRV
SLCDVACSLGNCAYFQLNAAIKSVHKTALDPSFAPNCVNFCLDKFLFRTLFLNGC"
72..3407
/codon_start=1
/product="126kDa protein"
/protein_id="BAB83987.1"
/db_xref="GI:18253268"
/translation="MAHQSTISNALLSVSGKNTLVNDLARRMYDFAVDFNARDR
RKVNPPKISSEOTLLVNAIPEQITPYNQAVHSLAGLRALELEYLMQVPGY
SPYIDGNNFAHLFGKRDYVHCMPNDLDIRMRHGGQDSIENYLSRLSRGKNVIP
EFQADIAAPAEVPCSCSTQFCQRIHPNPENSGRRYVAISLHSDYIPVHEFGAALI
SKDHH CYAAHFESEALLDQTEVLEIGATFKREGDDVGFPPADESTLVNKHKN
ILHYVVKSPVFPASSRIVVPKEPLVTRVNTWCKFTKVDYIILYKSVROVGCDSQFY
LADNPDFTYKALAMENTERAILRDTSVNFPPFKMOMVIVPLEFGSITSKQMTRESE
IVNRPDVTVLNHIQYQKALTYQNLVSFVESIKSRVLIINGVTARSEWDVKAILQP
LSMTFELQKLAALQDDIVMGKFCRLDKTISELWDEVGKFGNFPFTIKERLSRKI
LDVSENALKIKIPDYVTKORFAEYTKSEELPHLDIKDLEAEQMYDALSLSIL
RDADNFDIAKFMCKALCVNPDVAARVIVCAESGHLTNANLEHOLESLNDFHAKCV
ASEAVRLEPTSDVWVWVKFSIAERGLPVCAESGHLTNANLEHOLESLNDFHAKCV
DSVITKQWASVYVYTSGLVQOMKVVDSLAASLSATVSNLCKSLKDVVGYDSREKV
GYWDTLAKLQNLKTRMGREHEPTAKGVLVDGPGCGKTKKILERVDLDELILVPG
KQAAWIRRRANSGLIRATMDNVRTVDFLMHPKPTKRLFIIDEGMLHTGCVNPL
VLISGCDIAYIGDTQQIPTFNQVNPFPYKPFKLOVDEVERMTILRCFGDVPNFFL
QSKYEGAVSTTSVQORSVSEMIIGKGLNSVSKPLGKI VTFQADKFELEERKGYKN
VNTHEIQOETFPEDVSLVRLTATPLTLISKSSPHVLVALTRHTSKSPKYTVLVDPLVQ
IISDLSLSLSSFLLEMYVEAGGR"
4891..5661
/codon_start=1
/product="28kDa protein"
/protein_id="BAB83988.1"
/db_xref="GI:18253269"
/translation="MAVSLRDTVKISRFIDLSKODEILPAPMTKVKYRISTVDKIMA
VKNDSLVDLLKGVLVNGYVCLAGLVSGEWNLPDNCRGVSGVSVICVDRKMRSEK
ATLGAYHAPACKNFSFKALINYSITSEDAAKHPWQNLVNIKGVMEEGCPISLSEF
SICVYKRNKRGRLRLVTDGLPIELTEKVVVEEFVDEVPVAVKLERFRKTKRVV
GCVNNKKNINSKGLKVEIEDSVSDSESIASSSTP"
5667..6146
/codon_start=1
/product="17kDa protein"
/protein_id="BAB83989.1"
/db_xref="GI:18253270"
/translation="MPTIINSPSQFVYLSSAYADPVQILINLCTNALGNQPTQOARTT
VQOQFADAWKFPVPSITVRFASDFYVRYNSTDPLITALLINSIEDYRNIIEVDNPA
PNTTEIVNQRVDDATVATRASINNLANELVRGTGMFNQAGFETASGLVWTTTPAT"
6147..6146

ORIGIN
Query Match 91.6%; Score 69.6; DB 14; Length 6356;
Best Local Similarity 94.7%; Pred. No. 3.7e-13;
Matches 72; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GTGACAGACGGCTCGCCAAATGAACCTCACTGAAAGGTTGTCGAGGAGTTCGTGGATGA 60
Db 5431 GTAACAGACGGCTCGCCAAATGAACCTCACTGAAAGGTTGTCGAGGAGTTCGTGGATGA 5490
Qy 61 GTACCAATGGCTGTGA 76
Db 5491 GTACCAATGGCTGTGA 5506

RESULT 6
AX040184
LOCUS AX040184
DEFINITION Sequence 12 from Patent WO0063397.
ACCESSION AX040184
VERSION AX040184.1 GI:11230134
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
```

## REFERENCE

1 Meulwaeter, P., Cornelisse, Jacobs, J., van Eldik, G. and Metzlaaff, M. Methods and means for delivering inhibitory rna to plants and appl ications thereof  
Patent: WO 0063397-A 12 26-OCT-2000;  
Aventis CropScience N.V. (BE)  
Location/Qualifiers  
1..411  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="cDNA copy of part of the region of a TMV-U2 variant comprising the origin of assembly"

## FEATURES

source

## ORIGIN

Query Match 77.1%; Score 58.6; DB 6; Length 411;  
Best Local Similarity 93.8%; Pred. No. 1.4e-09;  
Matches 61; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
Qy 12 CTGCCCAAATTGAACCTCACTGAAAGGTTGTTGAGGAGTTCGTGGATGAAGTACCAATGGC 71  
Db 3 CTGCCCAAATTGAACCTCACTGAAAGGTTGTTGATGAGTTCTGATGATGAGTACCGATGGC 62  
Qy 72 TGTGA 76  
Db 63 TGTGA 67

## RESULT 7

ORVTRANS 912 bp RNA linear VRL 31-DEC-1990  
LOCUS Odontoglossum ringspot virus cell-to-cell transport gene.  
DEFINITION X55296  
ACCESSION X55296.1 GI:60833  
VERSION transport protein.  
KEYWORDS Odontoglossum ringspot virus  
SOURCE Odontoglossum ringspot virus  
ORGANISM Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.

## REFERENCE

1 Isomura, Y., Matumoto, Y., Murayama, A., Chatani, M., Inouye, N. and Ikegami, M.  
Nucleotide sequence of cell-to-cell transport protein gene of odontoglossum ringspot virus  
Nucleic Acids Res. 18 (24), 7448 (1990)  
91081343  
2259639  
2 (bases 1 to 912)  
Ikegami, M.  
Direct Submission  
Submitted (29-OCT-1990) Ikegami M., Nodai Research Institute, Tokyo  
University of Agriculture, 1-1-1 Sakuragaoka, Setagaya-ku, Tokyo  
156, JAPAN

## FEATURES

source

1..912  
/organism="Odontoglossum ringspot virus"  
/mol\_type="genomic RNA"  
/strain="Japan isolate"  
/db\_xref="taxon:12238"  
/clone="PORES-25"  
1..912  
/codon\_start=1  
/product="cell-to-cell transport protein"  
/protein\_id="CAA39008.1"  
/db\_xref="GI:60834"  
/db\_xref="GOA:P22590"  
/translation="MGRIRFVLLISFPIKTFSEPCSMALVLRDSIKISFINLSAS  
EKLPALTAIVKSVIRISKVDKISYEYNDLSDLLKGLKLVNGYVCLAGLVVYTGW  
NLPDNCRGVSGVSVICVDRKMRSEKATLSGYSACKRFTPKIPNTSVTTADALGKI  
WQVNTIRGVEMEKGFPLSLEFVSI CVVYLNINIKLGRKILNVTGGPTELTEAVI  
DEFVEKVPMAARKLSFRSVNKKPSNSSKFNKSLRNKLNLYENGSDSDVGSIVD  
DIVVNGVSDIRIDDCESPDQSDSY"

## CDS



```

mat_peptide 1..909
ORIGIN
  Query Match 60.3%; Score 45.8; DB 14; Length 912;
  Best Local Similarity 76.7%; Pred. No. 4.8e-05;
  Matches 56; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy 1 GTGACAGCGCTCGCCAAATGAACTCACTGAAAGGTTGTTGAGGAGTTCGTGGATGAA 60
Db 613 GTAACAGAGAGGCGCCACCGAACTTACTGAACAGTGTGTGATGAGTTCGTGGAGAA 672

Qy 61 GTACCAATGGCTG 73
Db 673 GTTCCTATGGCTG 685

RESULT 8
E03624
LOCUS
DEFINITION DNA encoding a part of complementary DNA to odontoglossum ringspot
virus genomic RNA.
ACCESSION E03624
VERSION E03624.1 GI:2171839
KEYWORDS JP 1992144685-A/1.
SOURCE Odontoglossum ringspot virus
ORGANISM Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
REFERENCE 1 (bases 1 to 5997)
AUTHORS Isomura,Y., Matsumoto,Y., Chatani,M. and Ikegami,M.
TITLES CDNA OF ORSV GENE
JOURNAL Patent: JP 1992144685-A 1 19-MAY-1992;
NIPPON OIL CO LTD
COMMENT OS odontoglossum ringspot virus
PN JP 1992144685-A/1
PD 19-MAY-1992
PF 28-NOV-1989 JP 1989306626
PI ISOMURA YOSHIKATSU, MATSUMOTO YOSHITOMO, CHATANI MASAOKI, PI
IKEGAMI MASATO
PC C12N15/40.C07K13/00.C07K15/04.C12N1/21.C12N5/10.C12P21/02, PC
(C12P21/02,
PC C12R1:19);
CC strandedness: Double;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
CC *source: clone=pORES-25;
FH Key Location/Qualifiers
FT misc_feature 1..5997
FT /note='a part of complementary DNA to FT
FT Odontoglossum
FT ringspot virus genomic RNA'
FT /note='a part of complementary DNA to ORSV'.
FT Location/Qualifiers
FEATURES
source
1..5997
/organism="Odontoglossum ringspot virus"
/mol_type="genomic RNA"
/db_xref="taxon:12238"

ORIGIN
  Query Match 60.3%; Score 45.8; DB 6; Length 5997;
  Best Local Similarity 76.7%; Pred. No. 7.1e-05;
  Matches 56; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy 1 GTGACAGCGCTCGCCAAATGAACTCACTGAAAGGTTGTTGAGGAGTTCGTGGATGAA 60
Db 4805 GTAACAGAGAGGCGCCACCGAACTTACTGAACAGTGTGTGATGAGTTCGTGGAGAA 4864

Qy 61 GTACCAATGGCTG 73
Db 4865 GTTCCTATGGCTG 4877

RESULT 10
ORU34586
LOCUS
DEFINITION Odontoglossum ringspot virus, complete genome.
ACCESSION U34586
VERSION U34586.1 GI:1407591
KEYWORDS Odontoglossum ringspot virus
SOURCE Odontoglossum ringspot virus
ORGANISM Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
REFERENCE 1 (bases 1 to 6609)
AUTHORS Chng,C.G., Wong,S.M., Mahtani,P.H., Loh,C.S., Goh,C.J., Kao,M.C.,
Chung,M.C. and Watanabe,Y.
TITLES The complete sequence of a Singapore isolate of odontoglossum

```

```

RESULT 9
E04305
LOCUS
DEFINITION CDNA encoding Odontoglossum ring spot virus (ORSV) genomic RNA.
ACCESSION E04305
VERSION E04305.1 GI:2172508
KEYWORDS JP 1993030975-A/1.
SOURCE Odontoglossum ringspot virus
ORGANISM Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
REFERENCE 1 (bases 1 to 6597)
AUTHORS Isomura,Y., Matsumoto,Y., Chatani,M., Mizuta,Y. and Ikegami,M.
TITLES CDNA OF ORSV GENE
JOURNAL Patent: JP 1993030975-A 1 09-FEB-1993;
NIPPON OIL CO LTD
COMMENT OS Odontoglossum ring spot virus
PN JP 1993030975-A/1
PD 09-FEB-1993
PF 26-JUL-1991 JP 1991276075
PI ISOMURA YOSHIKATSU, MATSUMOTO YOSHITOMO, CHATANI MASAOKI, PI
MIZUTA YOSHINORI, IKEGAMI MASATO
PC C12N15/40.C07K15/04.C12N1/21.C12N15/11.C12N15/70.C12P21/02, PC
C12Q1/68,
PC (C12N1/21.C12R1:19), (C12N15/70.C12R1:19), (C12P21/02.C12R1:19);
CC strandedness: Double;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
CC *source: clone=pORCP-D83;
FH Key Location/Qualifiers
FT misc_feature 1..6597
FT /note='cdna to Odontoglossum ring spot virus
FT genomic RNA'
FT CDS 5707..6183
FT /product='coat protein'
FT mat_peptide 5707..6180
FT /product='coat protein'.
FT Location/Qualifiers
source
1..6597
/organism="Odontoglossum ringspot virus"
/mol_type="genomic RNA"
/db_xref="taxon:12238"

ORIGIN
  Query Match 60.3%; Score 45.8; DB 6; Length 6597;
  Best Local Similarity 76.7%; Pred. No. 7.3e-05;
  Matches 56; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy 1 GTGACAGCGCTCGCCAAATGAACTCACTGAAAGGTTGTTGAGGAGTTCGTGGATGAA 60
Db 5405 GTAACAGAGAGGCGCCACCGAACTTACTGAACAGTGTGTGATGAGTTCGTGGAGAA 5464

Qy 61 GTACCAATGGCTG 73
Db 5465 GTTCCTATGGCTG 5477

RESULT 10
ORU34586
LOCUS
DEFINITION Odontoglossum ringspot virus, complete genome.
ACCESSION U34586
VERSION U34586.1 GI:1407591
KEYWORDS Odontoglossum ringspot virus
SOURCE Odontoglossum ringspot virus
ORGANISM Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
REFERENCE 1 (bases 1 to 6609)
AUTHORS Chng,C.G., Wong,S.M., Mahtani,P.H., Loh,C.S., Goh,C.J., Kao,M.C.,
Chung,M.C. and Watanabe,Y.
TITLES The complete sequence of a Singapore isolate of odontoglossum

```



DEAWBYKKTILAMLCBRTIIPDRAAVNFVFPKVMVIVLPDGSVTSGRMKRSEVMV  
NKDPVTVTLNHIITQDKALTYKNVLSFVESIRSVLINGSARSEWDVKSVLQALS  
MTFLQTLKAAQDQVLLKQFPDDTVTNLFWKQISDAVGDLPFSSIKETLISGGFVK  
VASQSLQITPDPBEEKIDPDVAVKVIATVIMQNELTISKPLERAKEYYNALSELVLKE  
CDEFDITQFNKLCEEKIDPDVAVKVIATVIMQNELTISKPLERAKEYYNALSELVLKE  
DMRFDLLKSLTCAAPPKSVKTLDSGLLPKQSGYDERQPSQSVVSDFHLKSVESVDM  
KMSAAVYTGPLKQVMKNYMDYLSASISATVSNLCKVLDKDVYGADEPESAKSGVYDV  
VKGKWLKPKGKCHANGVAELNNGEKVIVLLEWADGPPICGDMRRVAVSSDSLISDM  
GKLOTLLSKLQKDEPVSADKVTLDVDPGCGKTKKLETVNFNDEDLILVPGKEACM  
IIRKANKSGHVRATKDNVTVDSFLVHLKPKTKYNKLPIDEGMLHTGCVNFLTALSHC  
GLLENVAMTKRNFSPDLTGILDIETABLWNKFWDAYIIDLSESGNFTWMTSDAF  
HRWMAKQKSTIQLADFDVDFALPAIDQYKHMIAKQPKLDLSPODEYAAQTIVYH  
SKOINAFGLPSBELTQLLERIDSSKFLPYTRKTPEQIEBEFFSDLDSTVPMVEVLK  
ISKYAKSONEFHCAVEYLWEKLGNGFLSEVWQGRHRTSLKDYTAGITCWLAYORK  
SGDVTWPIGNVTIYIAACLASMIKPMVKVIAKAFCDGDSMLYIPKGLDLPDIQSGANLW  
NFAKLYRKRYGFCGYIITHDRGALVYTPDKLISGCKKHISLDHLEEFPRISLC  
DVSASLNCAYYQLNDATAIAEVHKTAVNGSFACFSIVKYLSDKNLFTLFLYNGSSTKG  
"

gene

CDS

/gene="126 kda protein"

66.. .3401

/gene="126 kda protein"

/codon start=1

/protein id="AAB49499.1"

/db xref="GI:1881787"

/translation="AHPQOTMNNKVIETAGMGRNSLINDLAQRVYDVAEELNHRSR  
PKYNSKVISQEQIIONAYPEFEITFYNTQLAVHSMAGGLRALELYLMMQIPGGS  
ITYDIGNFSAHKGKDYVHCMPNLDIRDVARHINQDVTSTYLARLERSKRGILP  
FOQSAFNKYNPDPAVCCDKRQDCSYSDLPKGYVALHSYDIPADBFGRALLR  
DVHICVAAFPSENLLETTSAPLDEIGATFYKSGDRLSFFQNESTLANYHESYKVI  
KYVCTPFPASNPVHKEPMCTRVNTWFKCTKVDTYFLFRGYTRGEDSEQFYTA  
DEAWBYKKTILAMLCBRTIIPDRAAVNFVFPKVMVIVLPDGSVTSGRMKRSEVMV  
NKDPVTVTLNHIITQDKALTYKNVLSFVESIRSVLINGSARSEWDVKSVLQALS  
MTFLQTLKAAQDQVLLKQFPDDTVTNLFWKQISDAVGDLPFSSIKETLISGGFVK  
VASQSLQITPDPBEEKIDPDVAVKVIATVIMQNELTISKPLERAKEYYNALSELVLKE  
CDEFDITQFNKLCEEKIDPDVAVKVIATVIMQNELTISKPLERAKEYYNALSELVLKE  
DMRFDLLKSLTCAAPPKSVKTLDSGLLPKQSGYDERQPSQSVVSDFHLKSVESVDM  
KMSAAVYTGPLKQVMKNYMDYLSASISATVSNLCKVLDKDVYGADEPESAKSGVYDV  
VKGKWLKPKGKCHANGVAELNNGEKVIVLLEWADGPPICGDMRRVAVSSDSLISDM  
GKLOTLLSKLQKDEPVSADKVTLDVDPGCGKTKKLETVNFNDEDLILVPGKEACM  
IIRKANKSGHVRATKDNVTVDSFLVHLKPKTKYNKLPIDEGMLHTGCVNFLTALSHC  
GLLENVAMTKRNFSPDLTGILDIETABLWNKFWDAYIIDLSESGNFTWMTSDAF  
HRWMAKQKSTIQLADFDVDFALPAIDQYKHMIAKQPKLDLSPODEYAAQTIVYH  
SKOINAFGLPSBELTQLLERIDSSKFLPYTRKTPEQIEBEFFSDLDSTVPMVEVLK  
ISKYAKSONEFHCAVEYLWEKLGNGFLSEVWQGRHRTSLKDYTAGITCWLAYORK  
SGDVTWPIGNVTIYIAACLASMIKPMVKVIAKAFCDGDSMLYIPKGLDLPDIQSGANLW  
NFAKLYRKRYGFCGYIITHDRGALVYTPDKLISGCKKHISLDHLEEFPRISLC  
DVSASLNCAYYQLNDATAIAEVHKTAVNGSFACFSIVKYLSDKNLFTLFLYNGSSTKG  
"

CDS

/gene="31 kda protein"

/codon start=1

/protein id="AAB49500.1"

/db xref="GI:1881788"

/translation="ALVIRDSIKISEFINISASEKLLPSALTAVKSVIRISKVKIISY  
ENPTLSDIDLKGLKGVYGVCLAGLVVTVGWNLPDNCCKGGVSLCLVDRMKRANSA  
TLGYSHTSACKQFTFKIIPNYSTYADALGKIWQVNTNIRGVEMEGECFPLSRKFS  
ICVYNNILKGLREKILNTVEGPTTELTAUVDEFEVKYPMARLKSFRSVNKKPFS  
NSKSFVNGKSRLSNRKNLANYENGSDVGISVDDIIVGVNGVSDIRIDDDCESPDQSD  
SY"

gene

CDS

/gene="31 kda protein"

5724.. .6197

/gene="18 kda protein"

5724.. .6197

/gene="18 kda protein"

/codon start=1

/protein id="AAB49501.1"

/db xref="GI:1881789"

/translation="SYTITDPSKLYLSSAWADPNLSINLTNSLGNQFOQOARTTV  
QQQPADVQVPTLTSRPPAGYFRVYVDPDILDLITLMTGTFEDFNRIILEVENPQ  
NPTTETLDRVDDATVAIRSAINLLNELLVNGTGMNQVQSPETISGLTWTSS"

## ORIGIN

Query Match 60.3%; Score 45.8; DB 14; Length 6611;  
Best Local Similarity 76.7%; Pred. No. 7.3e-05;  
Matches 56; Conservative 0; Mismatches 17; Indels 0; Gaps 0;  
QY 1 GTGACAGACGGCTCGCCAAATTGAACCTCAGTAAAGGTTCTGTCAGGAGTTTCGTGGATGAA 60  
DB 5419 GTAAACAGAGGAGGGCCACCGAATTACTTAAGACGAGTTGTTGATGAGTTTCGTGGAGAA 5478  
QY 61 GTACCAATGGCTG 73  
DB 5479 GTTCTCATGGCTG 5491

## RESULT 12

## AY571290

## LOCUS

Odontoglossum ringspot virus strain Taiwan, complete genome.

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## AUTHORS

## TITLE

## JOURNAL

## FEATURES

## source

## gene

## CDS

## protein

## translation

## translational

## translational

## translational

## translational

## translational

## translational

## translational

## translational

## translational

## translational

## translational

## translational

## translational

## translational

## translational

## translational

## translational

## translational

## translational

## translational

## translational

## translational

## translational

## translational

## translational

## translational

## translational

## translational

## translational

## translational

## translational

## translational

SAGLELVAMIKRNFSPDLTGILIEDTAELVNVKFWDAYIIDEISGNGVTMTSD  
AFHRMAQKSTTIGQLADFDVLPALDQYKHMKAQPKQKLDLSPODEYAAQTV  
YHSQINAFQLEFRQLERIDSSKFLFYTRKPEQEEFSDLSVPMVEVLE  
LDISKYQNEFHCAVEYLWEKLGNGFLEEVKMGQHRKTSKDTAGTKCLWQ  
RKSQDVTTFIGNTVIIAACLASMI PMDKVIAAFCGDDSIYIPKGLDLPDIQSGANL  
MNFPAKLYRKYGVFCGRYIIHHRGAIVVYDPLKLSKLGCKHLSLDHLEFRIS  
LCDYSSSLNACAYGQLNDIAEYHKTAVNGSPAFCSIVKYSKLNLFRTILFYNGSST  
KG"

CDS

64. .3402  
/gene="RdRp"  
/codon\_start=1  
/product="RNA-dependent RNA polymerase/replicase"  
/protein\_id="AAS87223.1"  
/db\_xref="GI:46309860"  
/translation="MAHFQQTMMNNKIVIEAGMRNSLINDLAQRVYDYNVAEELNHRSR  
RPKNFVSIVSQEQLIQATNAYPEFEITFYNTQAVHSMAGGLRALELEYLMQIPG  
SVTYIDIGNFQSAHLYKGRDYVHCMPNLDIRDVAHINQOQTVSTYTLARLERSKRGLP  
VFQOSAFNKMNDPDAVCCDKRFODCSYVDLPCKTYAVALHSIYDIPADEFGAALLR  
KDVIHCAAFHSENLLETTSAPLDEIGATFYKSGRLSFFQNESTLNTESYKNV  
IKYVCKTFFPASNRPVHKFEMCTRVNTFCFTKVDTYFLFRGVYTRGDSQFYTA  
MDEAWYKKTLLAMNSERTIPDRRAAVNFMFPKDMVIIVPLFDGVTSGMKRSEYM  
NKDFVTVTLNHIQYDKALTYKNVLSFVESIRSVIINGVTARSEWDVDSVLQAL  
SMTELOTKLAADQVVLKFKFPDPTVNLFWKOISDAVGLDLPSTIKERLISGGFV  
KVARSQIQTPTDEVIITADKLVMYKATELOHLDISKPLERAERYNALSELVYLK  
ECDFEDITQFNLCSEEDIDPDVYKVIIVAIMKNELDLPKNTPEALSALSPLPKD  
LDMFDELKUSTCAFFPSVKTLDSGLLPKQSGYDERQFESQSVVSVDPHLKSVEYK  
MKNWSSAVYTGPKVQOMKMYDLSASISATVSNLCKVLKVDYGDAPESAEKSGYVD  
VVGKWLKPKDKCHAWGKNGEYVILLEWADGFPICGDWRWRVAVSDSLIYSD  
MGKLTLLSKLDCGEPVSDAKVTLVDGPGCGKTKEILETVNDFEDLILVPGKEACK  
MIIRANKSGHVRATKQNRVTVDSFLMLKPKTYNKLFIDEGLMLHGTGCYNFLIALSH  
CREAMVGDTEQIPIINRVANFPYKHPATLYDHREVRRLSLKCPADVTHPMNSKYD  
GVNCTSDSVAEVRVGVGKPKSLKGLKLIITQSDKAELKTHGVEYEVSTFG  
EVNTVHRIQSTFEDVSVRLTPTPPLLELISKSSPHVLVALTRHTKSPKYYSVVLDPV  
KVCSDLKSVSDFLDMYKVDAGIL"

gene

CDS

4808. .5719  
/gene="Mp"  
/codon\_start=1  
/product="movement protein"  
/protein\_id="AAS87225.1"  
/db\_xref="GI:46309862"  
/translation="MGLRFRVLLSIPIKTFEPSPCTMALVLDSDIKISEFINLSAS  
EKLPSALTAVKSVRIKVDKILSYENDTLSDILLGKVLKLVNGYVCLAGLVVTGW  
NLPCNKGVSICLVDRKMRANEATIGSYHTSACKRFTFKIPNYSVTTADALRG  
WQMTNIRGVMEKGFCLSLFVSI CIVVNNIKLGRKILNVTGGGTTELTAAY  
DEFVKVPMAARLKSFRSVNKKPSNSSKFVNGKSLNSRNKLNYENGSDVGTSVVD  
DIVVNGVSDIRIDDDCESFDAQSDSY"  
5722. .6198  
/gene="Cp"  
/codon\_start=1  
/product="coat protein"  
/protein\_id="AAS87226.1"  
/db\_xref="GI:46309863"  
/translation="MSYTIIDPSKLYLSSAWADPNLSLNLCTNSLGNQFQTOQARTV  
VQOQADVWQFVFLTFRFPAGAGYFVRYDPLDPLITFLMGTFDTRNRIIEVENQAN  
QNPTTTETLDRVDDATVAIRSAINNLLNELVGRGTMYNQVSEFETMSGLTWTSS"

gene

CDS

5722. .6198  
/gene="Cp"  
/codon\_start=1  
/product="coat protein"  
/protein\_id="AAS87226.1"  
/db\_xref="GI:46309863"  
/translation="MSYTIIDPSKLYLSSAWADPNLSLNLCTNSLGNQFQTOQARTV  
VQOQADVWQFVFLTFRFPAGAGYFVRYDPLDPLITFLMGTFDTRNRIIEVENQAN  
QNPTTTETLDRVDDATVAIRSAINNLLNELVGRGTMYNQVSEFETMSGLTWTSS"

gene

CDS

5722. .6198  
/gene="Cp"  
/codon\_start=1  
/product="coat protein"  
/protein\_id="AAS87226.1"  
/db\_xref="GI:46309863"  
/translation="MSYTIIDPSKLYLSSAWADPNLSLNLCTNSLGNQFQTOQARTV  
VQOQADVWQFVFLTFRFPAGAGYFVRYDPLDPLITFLMGTFDTRNRIIEVENQAN  
QNPTTTETLDRVDDATVAIRSAINNLLNELVGRGTMYNQVSEFETMSGLTWTSS"

gene

CDS

5722. .6198  
/gene="Cp"  
/codon\_start=1  
/product="coat protein"  
/protein\_id="AAS87226.1"  
/db\_xref="GI:46309863"  
/translation="MSYTIIDPSKLYLSSAWADPNLSLNLCTNSLGNQFQTOQARTV  
VQOQADVWQFVFLTFRFPAGAGYFVRYDPLDPLITFLMGTFDTRNRIIEVENQAN  
QNPTTTETLDRVDDATVAIRSAINNLLNELVGRGTMYNQVSEFETMSGLTWTSS"

gene

CDS

5722. .6198  
/gene="Cp"  
/codon\_start=1  
/product="coat protein"  
/protein\_id="AAS87226.1"  
/db\_xref="GI:46309863"  
/translation="MSYTIIDPSKLYLSSAWADPNLSLNLCTNSLGNQFQTOQARTV  
VQOQADVWQFVFLTFRFPAGAGYFVRYDPLDPLITFLMGTFDTRNRIIEVENQAN  
QNPTTTETLDRVDDATVAIRSAINNLLNELVGRGTMYNQVSEFETMSGLTWTSS"

gene

CDS

5722. .6198  
/gene="Cp"  
/codon\_start=1  
/product="coat protein"  
/protein\_id="AAS87226.1"  
/db\_xref="GI:46309863"  
/translation="MSYTIIDPSKLYLSSAWADPNLSLNLCTNSLGNQFQTOQARTV  
VQOQADVWQFVFLTFRFPAGAGYFVRYDPLDPLITFLMGTFDTRNRIIEVENQAN  
QNPTTTETLDRVDDATVAIRSAINNLLNELVGRGTMYNQVSEFETMSGLTWTSS"

RESULT 13  
TOTW2  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
COMMENT  
FEATURES  
source  
1. .1004  
/organism="Tobacco mosaic virus"  
/mol\_type="genomic RNA"  
/db\_xref="taxon:12242"  
321. .800  
/note="unnamed protein product; reading frame (coat protein)"  
/codon\_start=1  
/protein\_id="CAA24695.1"  
/db\_xref="GI:62120"  
/db\_xref="GOA:P03570"  
/db\_xref="Swiss-Prot:P03570"  
/translation="MSYTIIDPSKLYLSSAWADPNLSLNLCTNSLGNQFQTOQARTV  
VQOQADVWQFVFLTFRFPAGAGYFVRYDPLDPLITFLMGTFDTRNRIIEVENQAN  
QNPTTTETLDRVDDATVAIRSAINNLLNELVGRGTMYNQVSEFETMSGLTWTSPAT"

Query Match 57.9%; Score 44; DB 14; Length 1004;  
Best Local Similarity 73.7%; Pred. No. 0.00021;  
Matches 56; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Qy 1 GTGACAGACGCTCGCCAAATGTCGA 76  
Db 52 GTGACAGACGCGGCGCCAACTACAGAAAGATTGTTGATGAGTTTCATGGAAGAT 111  
Qy 61 GTACCAATGGCTGTA 76  
Db 112 GTCCCTATGTCAATCA 127

ORIGIN  
Genomic RNA containing gene of capsid protein.  
E00088  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1 (bases 1 to 1019)  
Roorensu, I.P. and Merii, K.H.  
RNA PLANT VIRUS VECTOR OR FABRICATION AND UTILIZATION OF PART  
Patent: JP 1983051894-A 2 26-MAR-1983;  
NATL RES KAUNSRU OBU CANADA  
OS tobacco mosaic virus  
PN JP 1983051894-A/2  
PD 26-MAR-1983  
PF 27-MAY-1982 JP 1982090482  
PI ROORENSU II PERUCHIYAA, MERII KURISUTEIN HARASA PC  
C12N15/00, A01H1/00, C07H21/04, C12P19/34, C12P21/00//C12R1/91; CC  
strandedness: Single;

Query Match 57.9%; Score 44; DB 14; Length 1004;  
Best Local Similarity 73.7%; Pred. No. 0.00021;  
Matches 56; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Qy 1 GTGACAGACGCTCGCCAAATGTCGA 76  
Db 52 GTGACAGACGCGGCGCCAACTACAGAAAGATTGTTGATGAGTTTCATGGAAGAT 111  
Qy 61 GTACCAATGGCTGTA 76  
Db 112 GTCCCTATGTCAATCA 127

ORIGIN  
Genomic RNA containing gene of capsid protein.  
E00088  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1 (bases 1 to 1019)  
Roorensu, I.P. and Merii, K.H.  
RNA PLANT VIRUS VECTOR OR FABRICATION AND UTILIZATION OF PART  
Patent: JP 1983051894-A 2 26-MAR-1983;  
NATL RES KAUNSRU OBU CANADA  
OS tobacco mosaic virus  
PN JP 1983051894-A/2  
PD 26-MAR-1983  
PF 27-MAY-1982 JP 1982090482  
PI ROORENSU II PERUCHIYAA, MERII KURISUTEIN HARASA PC  
C12N15/00, A01H1/00, C07H21/04, C12P19/34, C12P21/00//C12R1/91; CC  
strandedness: Single;

Query Match 57.9%; Score 44; DB 14; Length 1004;  
Best Local Similarity 73.7%; Pred. No. 0.00021;  
Matches 56; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Qy 1 GTGACAGACGCTCGCCAAATGTCGA 76  
Db 52 GTGACAGACGCGGCGCCAACTACAGAAAGATTGTTGATGAGTTTCATGGAAGAT 111  
Qy 61 GTACCAATGGCTGTA 76  
Db 112 GTCCCTATGTCAATCA 127

ORIGIN  
Genomic RNA containing gene of capsid protein.  
E00088  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1 (bases 1 to 1019)  
Roorensu, I.P. and Merii, K.H.  
RNA PLANT VIRUS VECTOR OR FABRICATION AND UTILIZATION OF PART  
Patent: JP 1983051894-A 2 26-MAR-1983;  
NATL RES KAUNSRU OBU CANADA  
OS tobacco mosaic virus  
PN JP 1983051894-A/2  
PD 26-MAR-1983  
PF 27-MAY-1982 JP 1982090482  
PI ROORENSU II PERUCHIYAA, MERII KURISUTEIN HARASA PC  
C12N15/00, A01H1/00, C07H21/04, C12P19/34, C12P21/00//C12R1/91; CC  
strandedness: Single;

Query Match 57.9%; Score 44; DB 14; Length 1004;  
Best Local Similarity 73.7%; Pred. No. 0.00021;  
Matches 56; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Qy 1 GTGACAGACGCTCGCCAAATGTCGA 76  
Db 52 GTGACAGACGCGGCGCCAACTACAGAAAGATTGTTGATGAGTTTCATGGAAGAT 111  
Qy 61 GTACCAATGGCTGTA 76  
Db 112 GTCCCTATGTCAATCA 127

```

CC topology: linear;
CC hypothetical: No;
CC anti-sense: No; Location/Qualifiers
FH Key
FT 5'UTR 1..320
FT misc_feature 61..234
FT nuclear'/'note='region essential for formation of FT
FT misc_feature 235..320
FT 321..749 /note='region for regulation' FT CDS
FT 3'UTR 750..1022.
FEATURES
source
1..1019 Location/Qualifiers
/organism="Tobacco mosaic virus"
/mol_type="genomic RNA"
/db_xref="taxon:12242"
ORIGIN
Query Match 57.9%; Score 44; DB 6; Length 1019;
Best Local Similarity 73.7%; Pred. No. 0.00021;
Matches 56; Conservative 0; Mismatches 20; Indels 0; Gaps 0;
Qy 1 GTGACAGCGCTGCCAATGACTCACTGAAAGCTTTGTGAGGAGTTCGTGATGAA 60
Db 52 GTGACAGCGCGAGCGCCCTGGAAGCTTACAGAGAGCTTGTGATGAGTTCATGGAAGAT 111
Qy 61 GTACCAATGGCTGTGA 76
Db 112 GTCCCTATGTCATCA 127

RESULT 15
D63809
LOCUS D63809 6395 bp RNA linear VRL 13-FEB-1999
DEFINITION Tobacco mosaic virus genomic RNA for 130K protein, 180K protein,
30K protein and coat protein, complete sequence.
ACCESSION D63809.1 GI:1619995
VERSION coat protein; 30K protein; 180K protein; 130K protein.
KEYWORDS tobacco mosaic virus
SOURCE tobacco mosaic virus
ORGANISM
Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
Chen,J., Watanabe,Y., Sako,N., Ohshima,K. and Okada,Y.
Complete nucleotide sequence and synthesis of infectious in vitro
transcripts from a full-length cDNA clone of a rakkyo strain of
tobacco mosaic virus
JOURNAL Arch. Virol. 141 (5), 885-900 (1996)
MEDLINE 96265021
PUBMED 8678834
REFERENCE 2 (bases 1 to 6395)
AUTHORS Sako,N.
DIRECT SUBMISSION
TITLE Submitted (07-AUG-1995) Nobumichi Sako, Faculty of Agriculture,
Saga University, Laboratory of Plant Virology; 1 Honjo-machi, Saga,
Saga 840, Japan (Tel:0952-24-5191(ex.2730), Fax:0952-22-6274)
LOCATION/Qualifiers
1..6395
/organism="Tobacco mosaic virus"
/mol_type="genomic RNA"
/strain="Rakkyo"
/db_xref="taxon:12242"
69..4919
/note="position 3417..3419 is readthrough codon"
/codon_start=1
/transl_except=(pos:3417..3419,aa:OTHER)
/product="180K protein"
/protein_id="BAA09876.1"
/db_xref="GI:1619996"
/translacion="MAYTQTATTSALLDVTVRGNNSLVNDLAKRRLYDTAVDEFNARDR"
ORIGIN

```

```

RPKNFSKVISERQTLIATRAYPEFOITFYNTQNAVHSLAGLSRLSLELYLMMQIPY
SLTYDIGNPASHLFGRAVHCCMPNLDVRDMRHEGQKDSIELYLSRLDRGKGTVP
NFOKEAFDRYAEFPEADAVCHNTQTCEHQPMQOSGKYVATLHSHIVDIPADEFGAALL
RKVVHTCYAAPHFSENNLLEDSYVNLDEINACFSRQDGLTFTFASSTLNYCHFSFN
ILKYVCKTYFPASNRREYMKFVLTVNTWTFCKFSRIDTFLYKGVAAHKSVDSSQFYT
AMEDAWHYKKTLCAMCNSERILLEDSVSNVTFPPKMRDMVIIPLPDLSLETSSKRKEV
LVSKDQFVTVLHIRTVOAKALTYANLVSEFISRSRVINGVTARSEMDVDKSLLOS
LSMTFFLHTKLAVLKDDLLSKFSLGSKTVCQHVWDEISLAFGNAPPSVKERLANKL
IRAAGDALEIKVPDLVITFHDRLVAEYKSSVDMPALDIRKMEETEVMYNALSLSL
RESKDFDVFSQCKSLEVDPMTAAKVIVAVMSNESGLTITFPERPEANVAQALQDQ
EKASEGALVVTSEVEEPMKQMGAGELQAGFAGHPSSYSRNEEISLEQFHMA
TADSLIRKQMSIIVTGPVKQMQNFIDSLVASLSAAVNLVKILKDTAIDLETQ
KFGVLDVASKRWLIKPTAKSHANGVETHARKVHVALLEYDEQIVTCDWRRVAVSS
ESVYSDMAKLRTLRLRDGEPHVSNAKVLVLDVGVGCGKTKELLSRVNFDEDLILV
PGQAAMIRRRANSGLIVATKDNVTVDSFMMNFGKTRCQFRLFIDEGMLHTG
CNVFLVMSLCDVAYVIGDTQQIPIYNRVSGGFPYPAHFSLKLEGVDETRITLPCPAD
VTHYLNRRYEGFVSTSSVKSVSQEMVGAVINPISKPLKGLITFTQSDKEALLS
RGYSEVHTVEVQGETYSVSLVRLTPTPIISIIAGDSPHVLVALSRHTCSLKYYTVVM
DPLVSIIRDLKLSYLLDMYKVDAGIQ
69..3419
/codon_start=1
/product="130K protein"
/protein_id="BAA09877.1"
/db_xref="GI:1619997"
/translacion="MAYTQTATTSALLDVTVRGNNSLVNDLAKRRLYDTAVDEFNARDR
RPKNFSKVISERQTLIATRAYPEFOITFYNTQNAVHSLAGLSRLSLELYLMMQIPY
SLTYDIGNPASHLFGRAVHCCMPNLDVRDMRHEGQKDSIELYLSRLDRGKGTVP
NFOKEAFDRYAEFPEADAVCHNTQTCEHQPMQOSGKYVATLHSHIVDIPADEFGAALL
RKVVHTCYAAPHFSENNLLEDSYVNLDEINACFSRQDGLTFTFASSTLNYCHFSFN
ILKYVCKTYFPASNRREYMKFVLTVNTWTFCKFSRIDTFLYKGVAAHKSVDSSQFYT
AMEDAWHYKKTLCAMCNSERILLEDSVSNVTFPPKMRDMVIIPLPDLSLETSSKRKEV
LVSKDQFVTVLHIRTVOAKALTYANLVSEFISRSRVINGVTARSEMDVDKSLLOS
LSMTFFLHTKLAVLKDDLLSKFSLGSKTVCQHVWDEISLAFGNAPPSVKERLANKL
IRAAGDALEIKVPDLVITFHDRLVAEYKSSVDMPALDIRKMEETEVMYNALSLSL
RESKDFDVFSQCKSLEVDPMTAAKVIVAVMSNESGLTITFPERPEANVAQALQDQ
EKASEGALVVTSEVEEPMKQMGAGELQAGFAGHPSSYSRNEEISLEQFHMA
TADSLIRKQMSIIVTGPVKQMQNFIDSLVASLSAAVNLVKILKDTAIDLETQ
KFGVLDVASKRWLIKPTAKSHANGVETHARKVHVALLEYDEQIVTCDWRRVAVSS
ESVYSDMAKLRTLRLRDGEPHVSNAKVLVLDVGVGCGKTKELLSRVNFDEDLILV
PGQAAMIRRRANSGLIVATKDNVTVDSFMMNFGKTRCQFRLFIDEGMLHTG
CNVFLVMSLCDVAYVIGDTQQIPIYNRVSGGFPYPAHFSLKLEGVDETRITLPCPAD
VTHYLNRRYEGFVSTSSVKSVSQEMVGAVINPISKPLKGLITFTQSDKEALLS
RGYSEVHTVEVQGETYSVSLVRLTPTPIISIIAGDSPHVLVALSRHTCSLKYYTVVM
DPLVSIIRDLKLSYLLDMYKVDAGIQ"
4903..5709
/codon_start=1
/product="30K protein"
/protein_id="BAA09878.1"
/db_xref="GI:1619998"
/translacion="MALVVKGVNINEFIDTKMEKILPSMTFVKVSMVSKVDKIMV
HENESISEVNLKGVKIDSGYVCLAGLVVTGEWNLPCNCRGGVSVCLVDKMRERAD
ATLGSYYTAAKKRFQFQKVPVNYAITQDAMKNVQVLIINRVNMSAGFCPLSEFV
SCVIVYNNIKIGLRKTIYNRGDGPMEETVEVDFMEDVPMISIRLAKFRSRTGKKS
VVPKGNFSRDRSQPNKNYNAKDFGMSFKNNLIDDGSETSVAESDSF"
5712..6191
/codon_start=1
/product="coat protein"
/protein_id="BAA09879.1"
/db_xref="GI:1619999"
/translacion="MSYNTNPISQFVFLSSAWADPIELINLCTNALGNQFQTOQARTV
VQRFQSEVWKPSQVTVRFPDSDFKVYRFNVLDPVLTALIGAFDTNRNIIIEVNQAN
PSTAETILDTRRVDATVAIRSAINNLIIVELTRGTGYSNRSSPSSSSGLVMTSSPAT"

```

ORIGIN

Query Match 57.9%; Score 44; DB 14; Length 6395;  
Best Local Similarity 73.7%; Pred. No. 0.00031;  
Matches 56; Conservative 0; Mismatches 20; Indels 0; Gaps 0;  
Qy 1 GTGACAGACGGCTCGCCAAATTGAACTCACTGAAAAGGTTGTTGAGGAGTTCGTGGATCAA 60  
Db 5443 GTGAGACGCGGGGTCCCATGGAACTTACAGAGAGAGTTGTTGATGAGTTTCATCGAAGAT 5502  
Qy 61 GTACCAATGGCTGTGA 76  
Db 5503 GTACCTATGTCAATTA 5518

Search completed: January 17, 2005, 12:56:39  
Job time : 211.545 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 16, 2005, 21:45:12 ; Search time 22.7728 Seconds  
(without alignments)  
17518.955 Million cell updates/sec

Title: US-09-551-494-5\_COPY\_5430\_5505

Perfect score: 76

Sequence: 1 gtgacagagctgcgaat.....tgaagtaccatggtgtga 76

Scoring table: IDENTITY\_NUC

Gapop 10.0, Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq\_23Sep04:\*

- 1: Geneseqn1980s:\*
- 2: Geneseqn1990s:\*
- 3: Geneseqn2000s:\*
- 4: Geneseqn2001as:\*
- 5: Geneseqn2001bs:\*
- 6: Geneseqn2002as:\*
- 7: Geneseqn2002bs:\*
- 8: Geneseqn2003as:\*
- 9: Geneseqn2003bs:\*
- 10: Geneseqn2003cs:\*
- 11: Geneseqn2003ds:\*
- 12: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	76	100.0	6355	3 AAC62372	AAC62372 cDNA sequ
2	76	100.0	6355	9 ACC85005	Acc85005 TMV-U2 ge
3	69.6	91.6	769	10 ADI26338	Adi26338 Novel end
4	69.6	91.6	769	11 ADM68451	Adm68451 Tobacco m
5	69.6	91.6	769	12 ADP26604	Adp26604 Tobamovir
6	68	89.5	772	10 ADI26344	Adi26344 Novel end
7	68	89.5	772	11 ADM68457	Adm68457 Mosaic vi
8	68	89.5	772	12 ADP26610	Adp26610 Heterodup
9	58.6	77.1	411	3 AAC62379	AAC62379 Origin of
10	48.8	64.2	769	10 ADI26343	Adi26343 Novel end
11	48.8	64.2	769	11 ADM68456	Adm68456 Mosaic vi
12	48.8	64.2	769	12 ADP26609	Adp26609 Heterodup
13	45.8	60.3	5997	2 AAQ12188	Aaq12188 Odontoglo
14	45.8	60.3	6597	2 AAQ38106	Aaq38106 ORSV cDNA
15	44	57.9	356	1 AAN30116	Aan30116 TMV-RNA f
16	44	57.9	1019	1 AAN30115	Aan30115 TMV-RNA f
17	42.4	55.8	557	10 ADD17931	Add17931 DNA (Seqi
18	42.4	55.8	557	10 ADK56952	Adk56952 Plant DNA
19	42.4	55.8	632	2 AAV16847	Av16847 Tobacco m
20	42.4	55.8	792	2 ADI26342	Adi26342 Novel end
21	42.4	55.8	792	11 ADM68455	Adm68455 Mosaic vi

## ALIGNMENTS

RESULT 1  
AAC62372  
ID AAC62372 standard; DNA; 6355 BP.  
XX  
AC AAC62372;

DT 19-MAR-2001 (first entry)

DB cDNA sequence of the genome of tobacco mosaic virus-U2.

XX Plant phenotype; gene trait; Nicotiana; Oryza sativa; Zea mays; Brassica;  
KW Gossypium; Triticum; Arabidopsis; Petunia; herbicide; transgenic plant;  
KW tobacco necrosis virus; TMV; tobacco mosaic virus; TMV; helper virus; ss.  
XX Tobacco mosaic virus.

PN WO200063397-A2.

XX 26-OCT-2000.

XX 17-APR-2000; 2000WO-EP003521.

XX 20-APR-1999; 99US-00294022.

XX (AVET ) AVENTIS CROPS SCIENCE NV.

XX Meulawaeter F, Cornelissen M, Jacobs J, Van Eldik G, Metzlaiff M;

XX WPI; 2000-687182/67.

XX Identifying and isolating genes involved in determining the trait or  
PT phenotype of plant species, by infecting plants with gene silencing  
PT constructs targeted to the gene, and identifying plants with altered  
traits.

XX Example 1; Page 53-56; 64pp; English.

XX The specification describes a method for isolating genes that determine a  
CC trait or phenotype of a plant species. The method comprises identifying a  
CC set of nucleic acids of genes correlated with the trait, creating a  
CC library of gene silencing constructs in a viral RNA vector, targeting the  
CC gene silencing constructs to the nucleic acid set, infecting a collection  
CC of individual plants with these, identifying plants with altered traits  
CC or phenotype, and isolating genes of the invention. The method is useful

C 22 42.4 55.8 792 12 ADP26608  
23 42.4 55.8 804 12 ADM63297  
24 42.4 55.8 805 6 ABT06582  
25 42.4 55.8 807 6 ABT06572  
26 42.4 55.8 807 8 ABX14577  
27 42.4 55.8 807 11 ADM68440  
28 42.4 55.8 807 12 ADP26593  
29 42.4 55.8 895 2 AAO62653  
30 42.4 55.8 1187 10 ADD17932  
31 42.4 55.8 1187 10 ADK56953  
32 42.4 55.8 1825 2 AAQ67663  
33 42.4 55.8 1825 2 AAV54825  
34 42.4 55.8 1971 3 AAC62382  
35 42.4 55.8 5484 8 ADA15011  
36 42.4 55.8 6395 2 AAQ51555  
37 42.4 55.8 6395 2 AAZ20642  
38 42.4 55.8 6395 3 AAC62369  
39 42.4 55.8 6395 4 AAF82330  
40 42.4 55.8 6425 2 AAZ20644  
41 42.4 55.8 6425 4 AAF82332  
42 42.4 55.8 6439 2 AAZ20643  
43 42.4 55.8 6439 4 AAF82331  
44 42.4 55.8 6446 2 AAZ20646  
45 42.4 55.8 6446 4 AAF82334

Adp26608 Heterodup  
Adm63297 Tobamovir  
Abt06582 Wild-type  
Abt06572 Wild-type  
Abx14577 Tomato mo  
Adm68440 Tobacco m  
Adp26593 Tobamovir  
Aao62653 Tobacco-m  
Add17932 DNA (Seqi  
Adk56953 Plant DNA  
Aaq67663 TMV repli  
Aav54825 Replicon  
Aac62382 cDNA sequ  
Ada15011 Tobacco m  
Aaq51555 Tobacco m  
Aaz20642 TMV-based  
Aac62369 cDNA sequ  
Aaf82330 Wild-type  
Aaz20644 TMV-based  
Aaf82332 Tobacco m  
Aaz20643 TMV-based  
Aaf82331 Tobacco m  
Aaz20646 TMV-based  
Aaf82334 Tobacco m

for isolating genes involved in the determination of trait or a phenotype of a plant such as Nicotiana, Oryza sativa, Zea mays, Brassica, Gossypium, Triticum, Arabidopsis or Petunia. The method is also useful for modulating the expression of selected nucleic acid sequences and for validating the function of a nucleic acid sequence whose expression is correlated with the presence or absence of a specific trait in plants, but with otherwise unknown function. The method is also useful for developing agronomically useful products such herbicides or transgenic plants. The present sequence represents the cDNA sequence of the genome of tobacco mosaic virus (TMV)-U2. The sequence was used to construct a plasmid vector for the synthesis of an infective hybrid tobacco mosaic virus (TMV)/satellite tobacco necrosis virus (STNV) helper virus RNA. This helper virus is used in the method of the invention

XX SQ Sequence 6355 BP; 1933 A; 1112 C; 1489 G; 1821 T; 0 U; 0 Other;

Query Match 100.0%; Score 76; DB 3; Length 6355;  
Best Local Similarity 100.0%; Pred. No. 2.2e-16;  
Matches 76; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTGACAGCGCTCGCCCAATTGAACTCACTGAAAGGTTGTGAGGAGTTCTGTGGATGAA 60  
Db 5430 GTGACAGCGCTCGCCCAATTGAACTCACTGAAAGGTTGTGAGGAGTTCTGTGGATGAA 5489

Qy 61 GTACCAATGGCTGTGA 76  
Db 5490 GTACCAATGGCTGTGA 5505

RESULT 2  
ACC85005  
ID ACC85005 standard; DNA; 6355 BP.  
XX AC ACC85005;  
XX AC  
XX AC  
DT 13-OCT-2003 (first entry)  
XX  
DE TMV-U2 genome nucleotide sequence.  
XX  
XX Inhibitory RNA; viral RNA vector; coat protein; TMV; U2; gene; ds.

XX KW Tobacco mosaic virus.  
XX OS  
XX PN WO2003052108-A2.  
XX PD 26-JUN-2003.  
XX PF 05-DEC-2002; 2002WO-EP013964.  
XX PR 18-DEC-2001; 2001US-0340488P.  
XX PA (FARB ) BAYER BIOSCIENCE NV.  
XX PI Metzlaiff MH, Gossele VML, Meulewaeter F, Fache ICA;  
XX DR WPI; 2003-523529/49.  
XX  
XX Introducing inhibitory RNA into a plant cell comprises providing a viral RNA vector derived from a satellite RNA virus that encodes a coat protein, and infecting a plant with the viral RNA vector and a corresponding helper virus.

XX PS Example; Page 79-82; 86pp; English.  
XX  
XX The invention relates to introducing inhibitory RNA into a plant cell. The method involves providing a viral RNA vector derived from a satellite RNA virus having a sequence that encodes a coat protein, and infecting a plant with the viral RNA vector and a corresponding helper virus. The methods and viral RNA vectors are useful in introducing inhibitory RNA into plant cells. These may be used to determine or validate the function of isolated nucleic acid sequences in plants. The present sequence represents the nucleotide sequence of the genome of tobacco mosaic virus (TMV)-U2

XX SQ Sequence 6355 BP; 1933 A; 1112 C; 1489 G; 1821 T; 0 U; 0 Other;  
Query Match 100.0%; Score 76; DB 9; Length 6355;  
Best Local Similarity 100.0%; Pred. No. 2.2e-16;  
Matches 76; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTGACAGCGCTCGCCCAATTGAACTCACTGAAAGGTTGTGAGGAGTTCTGTGGATGAA 60  
Db 5430 GTGACAGCGCTCGCCCAATTGAACTCACTGAAAGGTTGTGAGGAGTTCTGTGGATGAA 5489

Qy 61 GTACCAATGGCTGTGA 76  
Db 5490 GTACCAATGGCTGTGA 5505

RESULT 3  
ADI26338  
ID ADI26338 standard; DNA; 769 BP.  
XX AC ADI26338;  
XX AC  
DT 22-APR-2004 (first entry)  
XX  
DE Novel endonuclease Res I-related clone DNA 3.  
XX  
XX endonuclease; molecular biology; plant propagation; phenotypic trait; herbicide tolerance; heat tolerance; cold tolerance; drought; salinity; osmotic stress; pest resistance; insect; nematode; arachnid; fungal; bacterial; viral; enzyme production; secondary metabolite; male sterility; female sterility; dwarfness; early maturity; Res I; ds.  
XX OS Tobacco mosaic virus.  
XX PN US2003148315-A1.  
XX PD 07-AUG-2003.  
XX PF 01-AUG-2002; 2002US-00211079.  
XX PR 01-FEB-2002; 2002US-0353722P.  
XX PR 14-MAR-2002; 2002US-00098155.  
XX PA (PADG/) PADGETT H S.  
XX PA (VAEW/) VAETHONGS A A.  
XX PI Padgett HS, Vaethongs AA;  
XX DR WPI; 2003-897548/82.  
XX  
XX New nucleic acid molecule encoding endonucleases, useful in molecular biology, specifically to generating populations of related nucleic acid molecules, and in plant propagation with useful phenotypic traits.  
XX PS Example 15; Fig 8; 46pp; English.  
XX  
XX This invention relates to a novel endonuclease (Res I) nucleic acid molecule which comprises a fully defined sequence of 899 bp given in the specification. The methods and compositions of the present invention are useful in molecular biology, and more specifically to generating populations of related nucleic acid molecules. They may also be used in plant propagation with useful phenotypic traits, such as improved tolerance to herbicides, improved tolerance to extremes of heat or cold, drought, salinity or osmotic stress, improved resistance to pests (insects, nematodes or arachnids) or diseases (fungal, bacterial or viral), production of enzymes or secondary metabolites, male or female sterility, dwarfness and early maturity. The present sequence is that of a clone which was derived during the exemplification of the invention.

XX SQ Sequence 769 BP; 247 A; 102 C; 201 G; 219 T; 0 U; 0 Other;  
Query Match 91.6%; Score 69.6; DB 10; Length 769;  
Best Local Similarity 94.7%; Pred. No. 2.3e-14;



Matches 72; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 GTGACAGCGCTCGCCAAATTGAACCTCACTGAAAAGTTGTTGAGGAGTTCTGGATGAA 60  
DB 541 GTAACAGACGCTCGCCAAATTGAACCTCACTGAAAAGTTGTTGAGGAGTTCTAGATGAA 600  
QY 61 GTACCAATGGCTGTGA 76  
DB 601 GTACCAATGGCTGTGA 616

RESULT 4  
ADM68451  
ID ADM68451 standard; DNA; 769 BP.  
XX  
AC ADM68451;  
DT 03-JUN-2004 (first entry)  
XX  
DE Tobacco mosaic virus movement protein gene #4.  
XX  
KW ds; mismatch endonuclease; endonuclease; gene shuffling technology;  
KW single nucleotide polymorphism; cancer susceptibility;  
KW sequence variation redistribution; movement protein; gene.  
XX  
OS Tobacco mosaic virus.  
XX  
PN US2003157682-A1.  
XX  
PD 21-AUG-2003.  
XX  
PF 31-JAN-2003; 2003US-00356708.  
XX  
PR 01-FEB-2002; 2002US-0353722P.  
PR 14-MAR-2002; 2002US-00098155.  
PR 01-AUG-2002; 2002US-00211079.  
XX  
PA (PADG/) PADGETT H S.  
PA (VAEW/) VAETHONGS A A.  
PA (VOJD/) VOJDANI F S.  
PA (SMIT/) SMITH M L.  
PA (LIND/) LINDBO J A.  
PA (FITZ/) FITZMAURICE W P.

XX  
PI Padgett HS, Vaethongs AA, Vojdani FS, Smith ML, Lindbo JA;  
PI Fitzmaurice WP;  
XX  
XX WPI; 2003-766176/72.  
XX  
XX Making a mismatch endonuclease, useful in gene shuffling and in detection  
PT of single nucleotide polymorphisms, comprises transfecting a host with a  
PT recombinant viral vector including a polynucleotide encoding a mismatch  
PT endonuclease.  
XX  
XX Example 14; SEQ ID NO 20; 79pp; English.

XX  
CC The invention relates to a method of making a mismatch endonuclease  
CC enzyme comprising transfecting a host plant, animal, yeast, fungus or  
CC bacterium with a recombinant viral vector that encodes a polynucleotide  
CC sequence for a mismatch endonuclease, growing the host so that the  
CC polynucleotide is expressed, and extracting the mismatch endonuclease  
CC enzyme from the host. The method is useful for making mismatch  
CC endonuclease enzymes, for obtaining peptides and polynucleotides with  
CC desired functional properties and for detecting mutations. The mismatch  
CC endonuclease enzymes are useful in gene shuffling technology for  
CC developing new genes, in detecting single nucleotide polymorphisms for  
CC e.g. detecting evidence of cancer susceptibility, or in redistributing  
CC sequence variations between non-identical polynucleotide sequences. The  
CC present sequence represents a tobacco mosaic virus movement protein gene.

XX  
SQ Sequence 769 BP; 247 A; 102 C; 201 G; 219 T; 0 U; 0 Other;  
Query Match 91.6%; Score 69.6; DB 11; Length 769;

Best Local Similarity 94.7%; Pred. No. 2.3e-14;  
Matches 72; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 GTGACAGCGCTCGCCAAATTGAACCTCACTGAAAAGTTGTTGAGGAGTTCTGGATGAA 60  
DB 541 GTAACAGACGCTCGCCAAATTGAACCTCACTGAAAAGTTGTTGAGGAGTTCTAGATGAA 600  
QY 61 GTACCAATGGCTGTGA 76  
DB 601 GTACCAATGGCTGTGA 616

RESULT 5  
ADP26604  
ID ADP26604 standard; DNA; 769 BP.  
XX  
AC ADP26604;  
DT 26-AUG-2004 (first entry)  
XX  
DE Tobamovirus DNA #5.

XX  
KW Sequence variation; heteroduplex; transcription; DNA integration;  
KW ribozyme expression; gene; ds; tobamovirus.

XX  
OS Tobamovirus.  
XX  
PN US2004110130-A1.  
XX  
PD 10-JUN-2004.  
XX  
PF 25-OCT-2002; 2002US-00280913.  
XX  
PR 02-FEB-2001; 2001US-0266386P.  
PR 14-FEB-2001; 2001US-0268785P.  
PR 01-FEB-2002; 2002US-00066390.  
PR 08-AUG-2002; 2002US-0402342P.  
XX  
PA (LARG-) LARGE SCALE BIOLOGY CORP.

XX  
PI Padgett HS, Lindbo JA, Fitzmaurice WP;  
XX  
XX WPI; 2004-440326/41.

XX  
PT Redistributing sequence variations between non-identical polynucleotide  
PT sequences, useful for generating improved polynucleotide having a desired  
PT characteristic, comprises making a heteroduplex and introducing a nick.

XX  
PS Example 15; SEQ ID NO 20; 75pp; English.

XX  
CC The invention relates to an in vitro method of redistributing sequence  
CC variations between non-identical polynucleotide sequences, comprising  
CC making a heteroduplex polynucleotide from two non-identical  
CC polynucleotides, introducing a nick in the second strand at or near a  
CC base pair mismatch site, removing the mismatched base(s) from the  
CC mismatch site where the nick occurred and using the first strand as a  
CC template to replace the removed base(s) with bases that complement the  
CC base(s) in the first strand. The invention also relates to an in vitro  
CC method of making a population of sequence variants from a heteroduplex  
CC polynucleotide sequence, obtaining a polynucleotide sequence encoding a  
CC desired functional property and identifying a reassorted DNA molecule  
CC encoding a protein with a desired functional property. The method is  
CC useful for generating an improved polynucleotide sequence or a population  
CC of improved polynucleotide sequences possessing at least one desired  
CC phenotypic characteristic (e.g., promotes transcription of linked  
CC polynucleotides), where such polynucleotides are useful for expression  
CC from a plant, animal, fungal, yeast, or bacterial expression vector, for  
CC integration to form a transgenic plant, animal or microorganism, and for  
CC expression of a ribozyme. This sequence represents DNA used in the scope  
CC of the invention.

XX  
SQ Sequence 769 BP; 247 A; 102 C; 201 G; 219 T; 0 U; 0 Other;

Query Match 91.6%; Score 69.6; DB 12; Length 769;  
 Best Local Similarity 94.7%; Pred. No. 2.3e-14;  
 Matches 72; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GTGACAGACGGCTCGCCAAATGAACTCACTGAAAAAGTTGTTGAGGAGTTCTGTGATGAA 60  
 |||||  
 Db 232 GTAAACAGACGGCTCGCCAAATGAACTCACTGAAAAAGTTGTTGAGGAGTTCTGTGATGAA 173  
 |||||

QY 61 GTACCAATGGCTGTGA 76  
 |||||  
 Db 172 GTACCAATGGCTGTGA 157  
 |||||

RESULT 7  
 ADM68457/c  
 ID ADM68457 standard; DNA; 772 BP.  
 XX  
 AC ADM68457;  
 XX  
 DT 03-JUN-2004 (first entry)  
 XX  
 DE Mosaic virus movement protein gene GRAMMR clone #11.  
 XX  
 KW ds; mismatch endonuclease; endonuclease; gene shuffling technology;  
 KW single nucleotide polymorphism; cancer susceptibility;  
 KW sequence variation redistribution; movement protein; gene.  
 XX  
 OS Tobacco mosaic virus.  
 OS Tomato mosaic virus.  
 XX  
 FN US2003157682-A1.  
 XX  
 PD 21-AUG-2003.  
 XX  
 PF 31-JAN-2003; 2003US-00356708.  
 XX  
 PR 01-FEB-2002; 2002US-0353722P.  
 PR 14-MAR-2002; 2002US-00098155.  
 PR 01-AUG-2002; 2002US-00211079.  
 XX  
 PA (PADG/) PADGETT H S.  
 PA (VAEW/) VAEWHONGS A A.  
 PA (VOJD/) VOJDANI F S.  
 PA (SMIT/) SMITH M L.  
 PA (LIND/) LINDBO J A.  
 PA (FITZ/) FITZMAURICE W P.  
 XX  
 PI Padgett HS, Vaewhongs AA, Vojdani FS, Smith ML, Lindbo JA;  
 PI Fitzmaurice WP;  
 XX  
 DR WPI; 2003-766176/72.  
 XX  
 PT Making a mismatch endonuclease, useful in gene shuffling and in detection  
 PT of single nucleotide polymorphisms, comprises transfecting a host with a  
 PT recombinant viral vector including a polynucleotide encoding a mismatch  
 PT endonuclease.  
 XX  
 PS Example 14; SEQ ID NO 26; 79pp; English.  
 XX  
 CC The invention relates to a method of making a mismatch endonuclease  
 CC enzyme comprising transfecting a host plant, animal, yeast, fungus or  
 CC bacterium with a recombinant viral vector that encodes a polynucleotide  
 CC sequence for a mismatch endonuclease, growing the host so that the  
 CC polynucleotide is expressed, and extracting the mismatch endonuclease  
 CC enzyme from the host. The method is useful for making mismatch  
 CC endonuclease enzymes, for obtaining peptides and polynucleotides with  
 CC desired functional properties and for detecting mutations. The mismatch  
 CC endonuclease enzymes are useful in gene shuffling technology for  
 CC developing new genes, in detecting single nucleotide polymorphisms for  
 CC e.g. detecting evidence of cancer susceptibility, or in redistributing  
 CC sequence variations between non-identical polynucleotide sequences. The  
 CC present sequence represents a mosaic virus movement protein gene GRAMMR  
 CC clone.  
 XX  
 SQ Sequence 772 BP; 227 A; 196 C; 109 G; 240 T; 0 U; 0 Other;  
 Query Match 89.5%; Score 68; DB 11; Length 772;

QY 1 GTGACAGACGGCTCGCCAAATGAACTCACTGAAAAAGTTGTTGAGGAGTTCTGTGATGAA 60  
 |||||  
 Db 541 GTAAACAGACGGCTCGCCAAATGAACTCACTGAAAAAGTTGTTGAGGAGTTCTGTGATGAA 600  
 |||||

QY 61 GTACCAATGGCTGTGA 76  
 |||||  
 Db 601 GTACCAATGGCTGTGA 616  
 |||||

RESULT 6  
 ADI26344/c  
 ID ADI26344 standard; DNA; 772 BP.  
 XX  
 AC ADI26344;  
 XX  
 DT 22-APR-2004 (first entry)  
 XX  
 DE Novel endonuclease Res I-related clone DNA 9.  
 XX  
 KW endonuclease; molecular biology; plant propagation; phenotypic trait;  
 KW herbicide tolerance; heat tolerance; cold tolerance; drought; salinity;  
 KW osmotic stress; pest resistance; insect; nematode; arachnid; fungal;  
 KW bacterial; viral; enzyme production; secondary metabolite;  
 KW male sterility; female sterility; dwarfness; early maturity; Res I; ds.  
 XX  
 OS Tobacco mosaic virus.  
 OS Tomato mosaic virus.  
 XX  
 FN US2003148315-A1.  
 XX  
 PD 07-AUG-2003.  
 XX  
 PF 01-AUG-2002; 2002US-00211079.  
 XX  
 PR 01-FEB-2002; 2002US-0353722P.  
 PR 14-MAR-2002; 2002US-00098155.  
 XX  
 PA (PADG/) PADGETT H S.  
 PA (VAEW/) VAEWHONGS A A.  
 XX  
 PI Padgett HS, Vaewhongs AA;  
 XX  
 DR WPI; 2003-897548/82.  
 XX  
 PT New nucleic acid molecule encoding endonucleases, useful in molecular  
 PT biology, specifically to generating populations of related nucleic acid  
 PT molecules, and in plant propagation with useful phenotypic traits.  
 XX  
 PS Example 15; Fig 14; 46pp; English.  
 XX  
 CC This invention relates to a novel endonuclease (Res I) nucleic acid  
 CC molecule which comprises a fully defined sequence of 899 bp given in the  
 CC specification. The methods and compositions of the present invention are  
 CC useful in molecular biology, and more specifically to generating  
 CC populations of related nucleic acid molecules. They may also be used in  
 CC plant propagation with useful phenotypic traits, such as improved  
 CC tolerance to herbicides, improved tolerance to extremes of heat or cold,  
 CC drought, salinity or osmotic stress, improved resistance to pests  
 CC (insects, nematodes or arachnids) or diseases (fungal, bacterial or  
 CC viral), production of enzymes or secondary metabolites, male or female  
 CC sterility, dwarfness and early maturity. The present sequence is that of  
 CC a clone which was derived during the exemplification of the invention.  
 XX  
 SQ Sequence 772 BP; 227 A; 196 C; 109 G; 240 T; 0 U; 0 Other;  
 Query Match 89.5%; Score 68; DB 10; Length 772;  
 Best Local Similarity 93.4%; Pred. No. 8.2e-14;  
 Matches 71; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Best Local Similarity 93.4%; Pred. No. 8.2e-14;  
Matches 71; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GTGACAGACGGCTCGCCAAATTGAACCTCACTGAAAGGTTGTTGAGGAGTTCGTGGATGAA 60  
|||  
Db 232 GTAACACAGCGCTCGCCAAATTGAACCTCACTGAAAGGTTGTTGATGATGTTTCATAGATGAA 173  
|||  
QY 61 GTACCAATGGCTGTGA 76  
|||  
Db 172 GTACCAATGGCTGTGA 157

## RESULT 8

ADP26610/c  
ID ADP26610 standard; DNA; 772 BP.

XX AC ADP26610;

XX DT 26-AUG-2004 (first entry)

XX DE Heteroduplex DNA #15.

XX Sequence variation; heteroduplex; transcription; DNA integration;

KW ribozyme expression; gene; ds.

XX OS Synthetic.

XX PN US2004110130-A1.

XX PD 10-JUN-2004.

XX PF 25-OCT-2002; 2002US-00280913.

XX PR 02-FEB-2001; 2001US-0266386P.

XX PR 14-FEB-2001; 2001US-0268785P.

XX PR 01-FEB-2002; 2002US-00066390.

XX PR 08-AUG-2002; 2002US-0402342P.

XX PA (LARG-) LARGE SCALE BIOLOGY CORP.

XX PI Padgett HS, Lindbo JA, Fitzmaurice WP;

XX WPI; 2004-440326/41.

XX Redistributing sequence variations between non-identical polynucleotide  
PT sequences, useful for generating improved polynucleotide having a desired  
PT characteristic, comprises making a heteroduplex and introducing a nick.

XX Example 15; SEQ ID NO 26; 75pp; English.

XX The invention relates to an in vitro method of redistributing sequence  
CC variations between non-identical polynucleotide sequences, comprising  
CC making a heteroduplex polynucleotide from two non-identical  
CC polynucleotides, introducing a nick in the second strand at or near a  
CC base pair mismatch site, removing the mismatched base(s) from the  
CC mismatch site where the nick occurred and using the first strand as a  
CC template to replace the removed base(s) with bases that complement the  
CC base(s) in the first strand. The invention also relates to an in vitro  
CC method of making a population of sequence variants from a heteroduplex  
CC polynucleotide sequence, obtaining a polynucleotide sequence encoding a  
CC desired functional property and identifying a reassorted DNA molecule  
CC encoding a protein with a desired functional property. The method is  
CC useful for generating an improved polynucleotide sequence or a population  
CC of improved polynucleotide sequences possessing at least one desired  
CC phenotypic characteristic (e.g., promotes transcription of linked  
CC polynucleotides), where such polynucleotides are useful for expression  
CC from a plant, animal, fungal, yeast, or bacterial expression vector, for  
CC integration to form a transgenic plant, animal or microorganism, and for  
CC expression of a ribozyme. This sequence represents DNA used in the scope  
CC of the invention.

XX Sequence 772 BP; 227 A; 196 C; 109 G; 240 T; 0 U; 0 Other;

Query Match 89.5%; Score 68; DB 12; Length 772;  
Best Local Similarity 93.4%; Pred. No. 8.2e-14;  
Matches 71; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GTGACAGACGGCTCGCCAAATTGAACCTCACTGAAAGGTTGTTGAGGAGTTCGTGGATGAA 60  
|||  
Db 232 GTAACACAGCGCTCGCCAAATTGAACCTCACTGAAAGGTTGTTGATGATGTTTCATAGATGAA 173  
|||

QY 61 GTACCAATGGCTGTGA 76  
|||  
Db 172 GTACCAATGGCTGTGA 157

## RESULT 9

AAC62379  
ID AAC62379 standard; DNA; 411 BP.

XX AC AAC62379;

XX DT 19-MAR-2001 (first entry)

XX DE Origin of assembly (OAS) of a tobacco mosaic virus (TMV)-U2.

XX Plant phenotype; gene trait; Nicotiana; Oryza sativa; Zea mays; Brassica;

KW Gossypium; Triticum; Arabidopsis; Petunia; herbicide; transgenic plant;

KW tobacco necrosis virus; TMV; tobacco mosaic virus; TMV; helper virus;

KW origin of assembly; ss.

XX OS Tobacco mosaic virus.

XX PN W0200063397-A2.

XX PD 26-OCT-2000.

XX PF 17-APR-2000; 2000WO-EP003521.

XX PR 20-APR-1999; 99US-00294022.

XX PA (AVET ) AVENTIS CROPS SCIENCE NV.

XX PI Meulewaeter F, Cornelissen M, Jacobs J, Van Eldik G, Metzlaiff M;

XX WPI; 2000-687182/67.

XX Identifying and isolating genes involved in determining the trait or  
PT phenotype of plant species, by infecting plants with gene silencing  
PT constructs targeted to the gene, and identifying plants with altered  
PT traits.

XX Example 1; Page 63; 64pp; English.

XX The specification describes a method for isolating genes that determine a  
CC trait or phenotype of a plant species. The method comprises identifying a  
CC set of nucleic acids of genes correlated with the trait, creating a  
CC library of gene silencing constructs in a viral RNA vector, targeting the  
CC gene silencing constructs to the nucleic acid set, infecting a collection  
CC of individual plants with these, identifying plants with altered traits  
CC or phenotype, and isolating genes of the invention. The method is useful  
CC for isolating genes involved in the determination of trait or a phenotype  
CC of a plant such as Nicotiana, Oryza sativa, Zea mays, Brassica, Gossypium,  
CC Triticum, Arabidopsis or Petunia. The method is also useful for  
CC modulating the expression of selected nucleic acid sequences and for  
CC validating the function of a nucleic acid sequence whose expression is  
CC correlated with the presence or absence of a specific trait in plants,  
CC but with otherwise unknown function. The method is also useful for  
CC developing agronomically useful products such herbicides or transgenic  
CC plants. The present sequence is an origin of assembly (OAS) of a tobacco  
CC mosaic virus (TMV)-U2. The sequence is used to construct infective hybrid  
CC tobacco mosaic virus (TMV)/tobacco necrosis virus (TMV) vectors, for use  
CC in the method of the invention

XX Sequence 411 BP; 140 A; 70 C; 93 G; 108 T; 0 U; 0 Other;

Query Match 77.1%; Score 58.6; DB 3; Length 411;  
 Best Local Similarity 93.8%; Pred. No. 1.4e-10;  
 Matches 61; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 12 CTCGCCAATTGAACCTCACTGAAAGGTTGTTGAGGAGTTCGTGGATGAAGTACCAATGGC 71  
 |||||  
 DB 3 CTCGCCAATTGAACCTCACTGAAAGGTTGTTGATGAGTTCGTGGATGAAGTACCAATGGC 62  
 |||||

QY 72 TGTGA 76  
 |||||  
 DB 63 TGTGA 67  
 |||||

RESULT 10  
 ADI26343/c  
 ID ADI26343 standard; DNA; 769 BP.  
 XX  
 AC ADI26343;  
 XX  
 DT 22-APR-2004 (first entry)  
 XX  
 DE Novel endonuclease Res I-related clone DNA 8.  
 XX  
 KW endonuclease; molecular biology; plant propagation; phenotypic trait;  
 KW herbicide tolerance; heat tolerance; cold tolerance; drought; salinity;  
 KW osmotic stress; pest resistance; insect; nematode; arachnid; fungal;  
 KW bacterial; viral; enzyme production; secondary metabolite;  
 KW male sterility; female sterility; dwarfness; early maturity; Res I; ds.  
 XX  
 OS Tobacco mosaic virus.  
 OS Tomato mosaic virus.  
 XX  
 US2003148315-A1.  
 XX  
 PD 07-AUG-2003.  
 XX  
 PF 01-AUG-2002; 2002US-00211079.  
 XX  
 PR 01-FEB-2002; 2002US-0353722P.  
 PR 14-MAR-2002; 2002US-00098155.  
 XX  
 PA (PADG/) PADGETT H S.  
 PA (VAEW/) VAEWHONGS A A.  
 XX  
 PI Padgett HS, Vaewhongs AA;  
 XX  
 WPI; 2003-897548/82.  
 XX  
 PT New nucleic acid molecule encoding endonucleases, useful in molecular  
 PT biology, specifically to generating populations of related nucleic acid  
 PT molecules, and in plant propagation with useful phenotypic traits.  
 XX  
 PS Example 15; Fig 13; 46pp; English.  
 XX  
 CC This invention relates to a novel endonuclease (Res I) nucleic acid  
 CC molecule which comprises a fully defined sequence of 899 bp given in the  
 CC specification. The methods and compositions of the present invention are  
 CC useful in molecular biology, and more specifically to generating  
 CC populations of related nucleic acid molecules. They may also be used in  
 CC plant propagation with useful phenotypic traits, such as improved  
 CC tolerance to herbicides, improved tolerance to extremes of heat or cold,  
 CC drought, salinity or osmotic stress, improved resistance to pests  
 CC (insects, nematodes or arachnids) or diseases (fungal, bacterial or  
 CC viral), production of enzymes or secondary metabolites, male or female  
 CC sterility, dwarfness and early maturity. The present sequence is that of  
 CC a clone which was derived during the exemplification of the invention.  
 XX  
 SQ Sequence 769 BP; 222 A; 205 C; 96 G; 246 T; 0 U; 0 Other;

Query Match 64.2%; Score 48.8; DB 10; Length 769;  
 Best Local Similarity 77.6%; Pred. No. 4.6e-07;  
 Matches 59; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 1 GTGACAGACGGCTCGCCAATTGAACCTCACTGAAAGGTTGTTGAGGAGTTCGTGGATGAA 60  
 |||||  
 DB 229 GTGAACGATGAGGACCCATGGAACCTTCAGAAGAAGTGTGATGAGTTCATAGATGAA 170  
 |||||

QY 61 GTACCAATGGCTGTGA 76  
 |||||  
 DB 169 GTACCAATGGCTGTGA 154  
 |||||

RESULT 11  
 ADM68456/c  
 ID ADM68456 standard; DNA; 769 BP.  
 XX  
 AC ADM68456;  
 XX  
 DT 03-JUN-2004 (first entry)  
 XX  
 DE Mosaic virus movement protein gene GRAMMR clone #10.  
 XX  
 KW ds; mismatch endonuclease; endonuclease; gene shuffling technology;  
 KW single nucleotide polymorphism; cancer susceptibility;  
 KW sequence variation redistribution; movement protein; gene.  
 XX  
 OS Tobacco mosaic virus.  
 OS Tomato mosaic virus.  
 XX  
 US2003157682-A1.  
 XX  
 PD 21-AUG-2003.  
 XX  
 PF 31-JAN-2003; 2003US-00356708.  
 XX  
 PR 01-FEB-2002; 2002US-0353722P.  
 PR 14-MAR-2002; 2002US-00098155.  
 PR 01-AUG-2002; 2002US-00211079.  
 XX  
 PA (PADG/) PADGETT H S.  
 PA (VAEW/) VAEWHONGS A A.  
 PA (VOJD/) VOJDANI F S.  
 PA (SMIT/) SMITH M L.  
 PA (LIND/) LINDBO J A.  
 PA (FITZ/) FITZMAURICE W P.  
 XX  
 PI Padgett HS, Vaewhongs AA, Vojdani FS, Smith ML, Lindbo JA;  
 PI Fitzmaurice WP;  
 XX  
 WPI; 2003-766176/72.  
 XX  
 PT Making a mismatch endonuclease, useful in gene shuffling and in detection  
 PT of single nucleotide polymorphisms, comprises transfecting a host with a  
 PT recombinant viral vector including a polynucleotide encoding a mismatch  
 PT endonuclease.  
 XX  
 PS Example 14; SEQ ID NO 25; 79pp; English.  
 XX  
 CC The invention relates to a method of making a mismatch endonuclease  
 CC enzyme comprising transfecting a host plant, animal, yeast, fungus or  
 CC bacterium with a recombinant viral vector that encodes a polynucleotide  
 CC sequence for a mismatch endonuclease, growing the host so that the  
 CC polynucleotide is expressed, and extracting the mismatch endonuclease  
 CC enzyme from the host. The method is useful for making mismatch  
 CC endonuclease enzymes, for obtaining peptides and polynucleotides with  
 CC desired functional properties and for detecting mutations. The mismatch  
 CC endonuclease enzymes are useful in gene shuffling technology for  
 CC developing new genes, in detecting single nucleotide polymorphisms for  
 CC e.g. detecting evidence of cancer susceptibility, or in redistributing  
 CC sequence variations between non-identical polynucleotide sequences. The  
 CC present sequence represents a mosaic virus movement protein gene GRAMMR  
 CC clone.  
 XX  
 SQ Sequence 769 BP; 222 A; 205 C; 96 G; 246 T; 0 U; 0 Other;

Query Match 64.2%; Score 48.8; DB 11; Length 769;



```
AC AAQ38106;
XX
XX 05-JUL-1993 (first entry)
XX
XX ORSV cDNA.
XX
XX Odontoglossum ring spot virus; screen; transformation; ds.
XX
XX Odontoglossum ring spot virus.
XX
XX JP05030975-A.
XX
XX 09-FEB-1993.
XX
XX 26-JUL-1991; 91JP-00276075.
XX
XX 26-JUL-1991; 91JP-00276075.
XX
XX (NIOC ) NIPPON OIL KK.
XX
XX WPI; 1993-087957/11.
XX
XX cDNA of odontoglossum ring-spot virus gene - useful as vector for plants
XX
XX Claim 2; Page 11-28; 30pp; Japanese.
XX
XX RNA was isolated from Odontoglossum ring spot virus and used to prepare
XX
XX cDNA. The DNA or its restriction fragments can be used to screen for ORSV
XX
XX or to detect genes related to ORSV. Vectors contg. the DNA sequence can
XX
XX be used to transform E. coli, Bacillus subtilis, Agrobacterium or plant
XX
XX cells for prodn. of the recombinant 130K, 180K and 30K proteins of ORSV,
XX
XX coat proteins or their peptide fragments
XX
XX Sequence 6597 BP; 1970 A; 1170 C; 1425 G; 2032 T; 0 U; 0 Other;
XX
XX Query Match 60.3%; Score 45.8; DB 2; Length 6597;
XX
XX Best Local Similarity 76.7%; Pred. NO. 9.2e-06;
XX
XX Matches 56; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
XX
XX Qy 1 GTGACAGACGCTCGCCAAATTGAACCTCACTGAAAGGTTGTGAGGAGTTCTGATGAA 60
XX
XX 5405 GTACAGAGGAGGCGCCACCGAATTACTGAAGCAGTTGTGTGAGTTCTGATGAA 5464
XX
XX 61 GTACCAATGGCTG 73
XX
XX 5465 GTTCCTATGGCTG 5477
XX
XX
XX RESULT 15
XX
XX ID AAN30116 standard; RNA; 356 BP.
XX
XX AC AAN30116;
XX
XX 25-MAR-2003 (revised)
XX
XX 02-NOV-1992 (first entry)
XX
XX TMV-RNA fragment I originating at the capped 5' end of the viral RNA and
XX
XX extending into the coat protein gene.
XX
XX RNA plant virus vector; tobacco mosaic virus; ss.
XX
XX Tobacco mosaic virus.
XX
XX Key Location/Qualifiers
XX
XX misc_feature 5
XX
XX /tag= e
XX
XX /label= bp No. 5400
XX
XX misc_feature 61..235
XX
XX /tag= a
XX
XX /label= nucleation region
XX
XX CDS 61..235
```

```
FT
FT misc_signal
FT 236..320
FT /tag= b
FT /label= control region
FT 348..356
FT /tag= d
FT /note= "site of cleavage by ribonuclease H"
XX
XX EP67553-A.
XX
XX 22-DEC-1982.
XX
XX 27-MAY-1981; 81US-00267539.
XX
XX 27-MAY-1981; 81US-00267539.
XX
XX (CANA ) NAT RES COUNCIL CANADA.
XX
XX Pelcher LE, Halasa MC;
XX
XX WPI; 1983-00323K/01.
XX
XX RNA plant virus vector from tobacco mosaic virus etc. - for modifying
XX
XX genes in plants to alter growth disease resistance etc.
XX
XX Example; Page 27; 56pp; English.
XX
XX The inventors claim an RNA plant virus vector from tobacco mosaic virus.
XX
XX The vector comprises a nucleotide sequence originating from the 5' end of
XX
XX the (-) strand of the viral RNA (Fragment I) (see AAN30116) and a
XX
XX sequence originating from the 3' end of the (-)strand (Fragment II) (see
XX
XX AAN30114, AAN30115). The RNA vector may have foreign genetic information
XX
XX inserted or attached, ultimately in the form of RNA, to the vector. The
XX
XX mode of Fragment I and II production fragments enables fragments of any
XX
XX desired length from any location in TMV-RNA to be generated (see
XX
XX AAN30117, AAN30118). Fragment I 76 is designated Frag.I CPG. The viral
XX
XX replicase gene is likely included Fragment I. (Updated on 25-MAR-2003 to
XX
XX correct PA field.)
XX
XX Qy Sequence 356 BP; 121 A; 48 C; 87 G; 0 T; 100 U; 0 Other;
XX
XX Query Match 57.9%; Score 44; DB 1; Length 356;
XX
XX Best Local Similarity 56.6%; Pred. No. 1.8e-05;
XX
XX Matches 43; Conservative 13; Mismatches 20; Indels 0; Gaps 0;
XX
XX Qy 1 GTGACAGACGCTCGCCAAATTGAACCTCACTGAAAGGTTGTGAGGAGTTCTGATGAA 60
XX
XX 52 GUGAGACGCGGAGGCGCCCAUGGAACUUCACAGAGAGUUGUUGAGUUGAUGAGAU 111
XX
XX Qy 61 GTACCAATGGCTGTGA 76
XX
XX 112 GUCCCUAUGUCAUCA 127
XX
XX Db 112 GUCCCUAUGUCAUCA 127
```

Search completed: January 17, 2005, 07:59:50  
Job time : 23.7728 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 17, 2005, 06:20:38 , Search time 4.08894 Seconds  
(without alignments)  
13211.223 Million cell updates/sec

Title: US-09-551-494-5\_COPY\_5430\_5505

Perfect score: 76

Sequence: 1 gtacacagagctcgcaat.....tgaagtacaaatggtgtga 76

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 824507 seqs, 35539441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:\*

- 1: /cgn2\_6/ptodata/1/ina/5A COMB.seq.\*
- 2: /cgn2\_6/ptodata/1/ina/5B COMB.seq.\*
- 3: /cgn2\_6/ptodata/1/ina/6A COMB.seq.\*
- 4: /cgn2\_6/ptodata/1/ina/6B COMB.seq.\*
- 5: /cgn2\_6/ptodata/1/ina/PCTUS COMB.seq.\*
- 6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42.4	55.8	433	1	US-07-971-101-5
2	42.4	55.8	807	4	US-09-565-616A-3
3	42.4	55.8	807	4	US-09-565-616A-4
4	42.4	55.8	1825	1	US-08-176-414B-1
5	42.4	55.8	1825	2	US-08-336-724-1
6	42.4	55.8	6395	2	US-08-687-559-2
7	42.4	55.8	6395	3	US-09-259-741-1
8	42.4	55.8	6395	3	US-09-037-751-1
9	42.4	55.8	6395	3	US-09-466-422-1
10	42.4	55.8	6395	4	US-03-401-415-2
11	42.4	55.8	6395	4	US-09-962-527-1
12	42.4	55.8	6425	3	US-09-259-741-3
13	42.4	55.8	6425	3	US-09-037-751-3
14	42.4	55.8	6425	3	US-09-466-422-3
15	42.4	55.8	6425	4	US-09-962-527-3
16	42.4	55.8	6439	3	US-09-259-741-2
17	42.4	55.8	6439	3	US-09-037-751-2
18	42.4	55.8	6439	3	US-09-466-422-2
19	42.4	55.8	6439	4	US-09-962-527-2
20	42.4	55.8	6446	3	US-09-259-741-5
21	42.4	55.8	6446	3	US-09-037-751-5
22	42.4	55.8	6446	3	US-09-466-422-5
23	42.4	55.8	6446	4	US-09-962-527-5
24	42.4	55.8	6475	3	US-09-259-741-4
25	42.4	55.8	6475	3	US-09-037-751-4
26	42.4	55.8	6475	3	US-09-466-422-4
27	42.4	55.8	6475	4	US-09-962-527-4

28 42.4 55.8 7685 3 US-09-502-710-22 Sequence 22, Appl  
29 42.4 55.8 7685 3 US-09-502-710-25 Sequence 25, Appl  
30 42.4 55.8 7685 3 US-09-502-711-22 Sequence 22, Appl  
31 42.4 55.8 7685 3 US-09-502-711-25 Sequence 25, Appl  
32 42.4 55.8 7685 4 US-09-565-616A-1 Sequence 1, Appl  
33 42.4 55.8 7686 3 US-09-502-710-23 Sequence 23, Appl  
34 42.4 55.8 7686 3 US-09-502-710-26 Sequence 26, Appl  
35 42.4 55.8 7686 3 US-09-502-711-23 Sequence 23, Appl  
36 42.4 55.8 7686 3 US-09-502-711-26 Sequence 26, Appl  
37 42.4 55.8 7686 4 US-09-565-616A-2 Sequence 2, Appl  
38 42.4 55.8 7687 3 US-09-502-710-24 Sequence 24, Appl  
39 42.4 55.8 7687 3 US-09-502-711-24 Sequence 24, Appl  
40 42.4 55.8 7688 3 US-09-502-710-27 Sequence 27, Appl  
41 42.4 55.8 7688 3 US-09-502-711-27 Sequence 27, Appl  
42 42.4 55.8 7926 3 US-09-500-554-1 Sequence 1, Appl  
43 42.4 55.8 7926 4 US-09-726-648-1 Sequence 1, Appl  
44 42.4 55.8 7926 4 US-10-119-330-1 Sequence 1, Appl  
45 41.4 54.5 75 1 US-07-971-101-6 Sequence 6, Appl

#### ALIGNMENTS

##### RESULT 1

US-07-971-101-5  
; Sequence 5, Application US/07971101  
; Patent No. 543969  
; GENERAL INFORMATION:  
; APPLICANT: Wilson, Thomas M.A. et al.  
; TITLE OF INVENTION: RNA Packaging System  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/971,101  
; FILING DATE: 19921029  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 7108-006  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212 790-9090  
; TELEFAX: 212 869-9741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 433 nucleotides  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: single  
; TOPOLOGY: unknown  
; MOLECULE TYPE: RNA  
US-07-971-101-5

Query Match 55.8%; Score 42.4; DB 1; Length 433;  
Best Local Similarity 56.6%; Pred. No. 1.6e-06;  
Matches 43; Conservative 12; Mismatches 21; Indels 0; Gaps 0;

Qy 1 GTGACAGCGCTCGCCCAATTGAACCTCACTGAAAGTTGTTGAGGAGTTCGTGATGAA 60

Db 326 GUGAGAGACGGGGGCCCAUGGAACUACAGAGAGUCGUGAUGAUGAUGAUGAUGAUGA 385

Qy 61 GTACCAATGGCTGTGA 76





```
;
;
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/336,724
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/997,733
; FILING DATE: 30-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Halluin, Albert P.
; REGISTRATION NUMBER: 25,227
; REFERENCE/DOCKET NUMBER: BIOG-20220 USA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-433-4150
; TELEFAX: 415-433-8716
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1825
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: RNA (episomal), peptide
; DESCRIPTION: Peptide encodes for TMV 30kDa
; DESCRIPTION: movement protein (268 residues) and CAT (204
; DESCRIPTION: residues).
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Tobacco Mosaic Virus
; IMMEDIATE SOURCE:
; CLONE:
; FEATURE:
;
; US-08-336-724-1
;
; Query Match 55.8%; Score 42.4; DB 2; Length 1825;
; Best Local Similarity 56.6%; Pred. No. 2.7e-06;
; Matches 43; Conservative 12; Mismatches 21; Indels 0; Gaps 0;
;
; QY 1 GTGACAGCGCTCGCCAAATGAACTCACTGAAAGGTTGTGAGGAGTTCGTGGAGAA 60
; Db 610 GUGAGAGCGAGGCGGCCCGAAGGAACTGAGGAGGAGGAGGAGGAGGAGGAGG 669
;
; QY 61 GTACCAATGGCTGTGA 76
; Db 670 GUCCUAUGUGCAUCA 685
;
; RESULT 6
; US-08-687-559-2
; Sequence 2, Application US/08687559
; Patent No. 5955647
; GENERAL INFORMATION:
; APPLICANT: Fitchen, John H.
; APPLICANT: Beachy, Roger N.
; TITLE OF INVENTION: METHOD FOR USING TOBACCO MOSAIC VIRUS TO
; TITLE OF INVENTION: OVERPRODUCE PEPTIDES AND PROTEINS
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
;
; US-08-687-559-2
; Sequence 1, Application US/09259741
; Patent No. 6031895
; GENERAL INFORMATION:
; APPLICANT: GARGER, STEPHEN
; APPLICANT: HOLTZ, R. BARRY
; APPLICANT: McCULLOCH, MICHAEL
; APPLICANT: TURPEN, THOMAS
; TITLE OF INVENTION: A PROCESS FOR ISOLATING AND
; TITLE OF INVENTION: PURIFYING VIRUSES SOLUBLE PROTEINS AND PEPTIDES FROM PLANT
; TITLE OF INVENTION: SOURCES
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howrey & Simon
; STREET: 1299 Pennsylvania Avenue N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/259,741
; FILING DATE: February 25, 1999
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/037,751
; FILING DATE: March 10, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Halluin, Albert P
```

```
;
;
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,559
; FILING DATE: No. 5955647ember 18, 1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/01467
; FILING DATE: 03-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Learn, June M.
; REGISTRATION NUMBER: 31,238
; REFERENCE/DOCKET NUMBER: 07302/011001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-678-5070
; TELEFAX: 619-678-5099
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6395 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; IMMEDIATE SOURCE:
; CLONE: TMV
;
; US-08-687-559-2
;
; Query Match 55.8%; Score 42.4; DB 2; Length 6395;
; Best Local Similarity 72.4%; Pred. No. 4.3e-06;
; Matches 55; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
;
; QY 1 GTGACAGCGCTCGCCAAATGAACTCACTGAAAGGTTGTGAGGAGTTCGTGGAGAA 60
; Db 5443 GTGAGAGCGAGGCGGCCCGAAGGAACTTACAGAGAGCTCGTTGATGATTCGGAAGAT 5502
;
; QY 61 GTACCAATGGCTGTGA 76
; Db 5503 GTCCCTATGTCGATCA 5518
;
; RESULT 7
; US-09-259-741-1
; Sequence 1, Application US/09259741
; Patent No. 6031895
; GENERAL INFORMATION:
; APPLICANT: GARGER, STEPHEN
; APPLICANT: HOLTZ, R. BARRY
; APPLICANT: McCULLOCH, MICHAEL
; APPLICANT: TURPEN, THOMAS
; TITLE OF INVENTION: A PROCESS FOR ISOLATING AND
; TITLE OF INVENTION: PURIFYING VIRUSES SOLUBLE PROTEINS AND PEPTIDES FROM PLANT
; TITLE OF INVENTION: SOURCES
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howrey & Simon
; STREET: 1299 Pennsylvania Avenue N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/259,741
; FILING DATE: February 25, 1999
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/037,751
; FILING DATE: March 10, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Halluin, Albert P
```

	REGISTRATION NUMBER:	25,277	
	REFERENCE/DOCKET NUMBER:	00801.0140.US01	
	TELECOMMUNICATION INFORMATION:		
	TELEPHONE:	650-463-8100	
	TELEFAX:	650-463-8400	
	TELEX:		
	INFORMATION FOR SEQ ID NO: 1:		
	SEQUENCE CHARACTERISTICS:		
	LENGTH:	6395 base pairs	
	TYPE:	nucleic acid	
	STRANDEDNESS:	single	
	TOPOLOGY:	unknown	
	MOLECULE TYPE:	Genomic RNA	
	US-09-259-741-1		
	Query Match	55.8%; Score 42.4; DB 3; Length 6395;	
	Best Local Similarity	56.6%; Pred. No. 4.3e-06;	
	Matches	43; Conservative 12; Mismatches 21; Indels 0; Gaps 0;	
QY	1 GTGACAGACGGCTGCCCAATTGAACACTCAGTAAAGGTGTTGAGGAGTTCTGTGGATGAA 60		
DB	5443 GUGAGAGACGGAGCGCCCAUGAACUACAGAGAAGUCGUUGAUGAUCAUGGAAGAU 5502		
QY	61 GTACCAATGGCTGTGA 76		
DB	5503 GUCCCUAUGUCAUCA 5518		
	RESULT 8		
	US-09-037-751-1		
	; Sequence 1, Application US/09037751		
	; Patent No. 6037456		
	; GENERAL INFORMATION:		
	; APPLICANT: GARGER, STEPHEN		
	; APPLICANT: HOLTZ, R. BARRY		
	; APPLICANT: MCCULLOCH, MICHAEL		
	; APPLICANT: TURPEN, THOMAS		
	; TITLE OF INVENTION: A PROCESS FOR ISOLATING AND PURIFYING VIRUSES SOLUBLE PROTEINS AND PEPTIDES		
	; FROM PLANT SOURCES		
	; NUMBER OF SEQUENCES: 5		
	; CORRESPONDENCE ADDRESS:		
	; ADDRESSEE: Howrey & Simon		
	; STREET: 1299 Pennsylvania Avenue N.W.		
	; CITY: Washington		
	; STATE: DC		
	; COUNTRY: USA		
	; ZIP: 20004		
	; COMPUTER READABLE FORM:		
	; MEDIUM TYPE: Diskette		
	; COMPUTER: IBM Compatible		
	; OPERATING SYSTEM: DOS		
	; SOFTWARE: FastSeq for Windows Version 2.0		
	; CURRENT APPLICATION DATA:		
	; APPLICATION NUMBER: US/09/037,751		
	; FILING DATE: 10-MAR-1998		
	; CLASSIFICATION:		
	; PRIOR APPLICATION DATA:		
	; APPLICATION NUMBER:		
	; FILING DATE:		
	; ATTORNEY/AGENT INFORMATION:		
	; NAME: Halluin, Albert P		
	; REGISTRATION NUMBER: 25,277		
	; REFERENCE/DOCKET NUMBER: 00801.0140.999		
	; TELECOMMUNICATION INFORMATION:		
	; TELEPHONE: 650-463-8109		
	; TELEFAX: 650-463-8400		
	; TELEX:		
	; INFORMATION FOR SEQ ID NO: 1:		
	SEQUENCE CHARACTERISTICS:		
	LENGTH: 6395 base pairs		
	TYPE: nucleic acid		
	STRANDEDNESS: single		
	TOPOLOGY: unknown		
	MOLECULE TYPE: Genomic RNA		
	SEQUENCE DESCRIPTION: SEQ ID NO: 1:		
	US-09-466-422-1		
	Query Match	55.8%; Score 42.4; DB 3; Length 6395;	
	Best Local Similarity	56.6%; Pred. No. 4.3e-06;	
	Matches	43; Conservative 12; Mismatches 21; Indels 0; Gaps 0;	
QY	1 GTGACAGACGGCTGCCCAATTGAACACTCAGTAAAGGTGTTGAGGAGTTCTGTGGATGAA 60		
DB	5443 GUGAGAGACGGAGCGCCCAUGAACUACAGAGAAGUCGUUGAUGAUCAUGGAAGAU 5502		
QY	61 GTACCAATGGCTGTGA 76		
DB	5503 GUCCCUAUGUCAUCA 5518		
	RESULT 9		
	US-09-466-422-1		
	; Sequence 1, Application US/09466422		
	; Patent No. 6303779		
	; GENERAL INFORMATION:		
	; APPLICANT: GARGER, STEPHEN		
	; APPLICANT: HOLTZ, R. BARRY		
	; APPLICANT: MCCULLOCH, MICHAEL		
	; APPLICANT: TURPEN, THOMAS		
	; TITLE OF INVENTION: A PROCESS FOR ISOLATING AND PURIFYING VIRUSES SOLUBLE PROTEINS AND PEPTIDES		
	; FROM PLANT SOURCES		
	; NUMBER OF SEQUENCES: 5		
	; CORRESPONDENCE ADDRESS:		
	; ADDRESSEE: Howrey & Simon		
	; STREET: 1299 Pennsylvania Avenue N.W.		
	; CITY: Washington		
	; STATE: DC		
	; COUNTRY: USA		
	; ZIP: 20004		
	; COMPUTER READABLE FORM:		
	; MEDIUM TYPE: Diskette		
	; COMPUTER: IBM Compatible		
	; OPERATING SYSTEM: DOS		
	; SOFTWARE: FastSeq for Windows Version 2.0		
	; CURRENT APPLICATION DATA:		
	; APPLICATION NUMBER: US/09/037,751		
	; FILING DATE: 17-Dec-1999		
	; CLASSIFICATION: <Unknown>		
	; PRIOR APPLICATION DATA:		
	; APPLICATION NUMBER: 09/037,751		
	; FILING DATE: <Unknown>		
	; ATTORNEY/AGENT INFORMATION:		
	; NAME: Halluin, Albert P		
	; REGISTRATION NUMBER: 25,277		
	; REFERENCE/DOCKET NUMBER: 00801.0140.999		
	; TELECOMMUNICATION INFORMATION:		
	; TELEPHONE: 650-463-8109		
	; TELEFAX: 650-463-8400		
	; TELEX: <Unknown>		
	; INFORMATION FOR SEQ ID NO: 1:		
	SEQUENCE CHARACTERISTICS:		
	LENGTH: 6395 base pairs		
	TYPE: nucleic acid		
	STRANDEDNESS: single		
	TOPOLOGY: unknown		
	MOLECULE TYPE: Genomic RNA		
	SEQUENCE DESCRIPTION: SEQ ID NO: 1:		
	US-09-466-422-1		
	Query Match	55.8%; Score 42.4; DB 3; Length 63	

	REGISTRATION NUMBER:	25,277	
	REFERENCE/DOCKET NUMBER:	00801.0140.US01	
	TELECOMMUNICATION INFORMATION:		
	TELEPHONE:	650-463-8100	
	TELEFAX:	650-463-8400	
	TELEX:		
	INFORMATION FOR SEQ ID NO: 1:		
	SEQUENCE CHARACTERISTICS:		
	LENGTH:	6395 base pairs	
	TYPE:	nucleic acid	
	STRANDEDNESS:	single	
	TOPOLOGY:	unknown	
	MOLECULE TYPE:	Genomic RNA	
	US-09-259-741-1		
	Query Match	55.8%; Score 42.4; DB 3; Length 6395;	
	Best Local Similarity	56.6%; Pred. No. 4.3e-06;	
	Matches	43; Conservative 12; Mismatches 21; Indels 0; Gaps 0;	
QY	1 GTGACAGACGGCTGCCCAATTGAACCTCAGTAAAGGTGTTGAGGAGTTCTGTGGATGAA 60		
DB	5443 GUGAGAGACGGAGCGCCCAUGAACUACAGAGAAGUCGUUGAUGAUCAUGGAAGAU 5502		
QY	61 GTACCAATGGCTGTGA 76		
DB	5503 GUCCCUAUGUCAUCA 5518		
	RESULT 8		
	US-09-037-751-1		
	; Sequence 1, Application US/09037751		
	; Patent No. 6037456		
	; GENERAL INFORMATION:		
	; APPLICANT: GARGER, STEPHEN		
	; APPLICANT: HOLTZ, R. BARRY		
	; APPLICANT: MCCULLOCH, MICHAEL		
	; APPLICANT: TURPEN, THOMAS		
	; TITLE OF INVENTION: A PROCESS FOR ISOLATING AND PURIFYING VIRUSES SOLUBLE PROTEINS AND PEPTIDES		
	; FROM PLANT SOURCES		
	; NUMBER OF SEQUENCES: 5		
	; CORRESPONDENCE ADDRESS:		
	; ADDRESSEE: Howrey & Simon		
	; STREET: 1299 Pennsylvania Avenue N.W.		
	; CITY: Washington		
	; STATE: DC		
	; COUNTRY: USA		
	; ZIP: 20004		
	; COMPUTER READABLE FORM:		
	; MEDIUM TYPE: Diskette		
	; COMPUTER: IBM Compatible		
	; OPERATING SYSTEM: DOS		
	; SOFTWARE: FastSeq for Windows Version 2.0		
	; CURRENT APPLICATION DATA:		
	; APPLICATION NUMBER: US/09/037,751		
	; FILING DATE: 10-MAR-1998		
	; CLASSIFICATION:		
	; PRIOR APPLICATION DATA:		
	; APPLICATION NUMBER:		
	; FILING DATE:		
	; ATTORNEY/AGENT INFORMATION:		
	; NAME: Halluin, Albert P		
	; REGISTRATION NUMBER: 25,277		
	; REFERENCE/DOCKET NUMBER: 00801.0140.999		
	; TELECOMMUNICATION INFORMATION:		
	; TELEPHONE: 650-463-8109		
	; TELEFAX: 650-463-8400		
	; TELEX:		
	; INFORMATION FOR SEQ ID NO: 1:		
	SEQUENCE CHARACTERISTICS:		
	LENGTH: 6395 base pairs		
	TYPE: nucleic acid		
	STRANDEDNESS: single		
	TOPOLOGY: unknown		
	MOLECULE TYPE: Genomic RNA		
	SEQUENCE DESCRIPTION: SEQ ID NO: 1:		
	US-09-466-422-1		
	Query Match	55.8%; Score 42.4; DB 3; Length 6395;	
	Best Local Similarity	56.6%; Pred. No. 4.3e-06;	
	Matches	43; Conservative 12; Mismatches 21; Indels 0; Gaps 0;	
QY	1 GTGACAGACGGCTGCCCAATTGAACCTCAGTAAAGGTGTTGAGGAGTTCTGTGGATGAA 60		
DB	5443 GUGAGAGACGGAGCGCCCAUGAACUACAGAGAAGUCGUUGAUGAUCAUGGAAGAU 5502		
QY	61 GTACCAATGGCTGTGA 76		
DB	5503 GUCCCUAUGUCAUCA 5518		
	RESULT 9		
	US-09-466-422-1		
	; Sequence 1, Application US/09466422		
	; Patent No. 6303779		
	; GENERAL INFORMATION:		
	; APPLICANT: GARGER, STEPHEN		
	; APPLICANT: HOLTZ, R. BARRY		
	; APPLICANT: MCCULLOCH, MICHAEL		
	; APPLICANT: TURPEN, THOMAS		
	; TITLE OF INVENTION: A PROCESS FOR ISOLATING AND PURIFYING VIRUSES SOLUBLE PROTEINS AND PEPTIDES		
	; FROM PLANT SOURCES		
	; NUMBER OF SEQUENCES: 5		
	; CORRESPONDENCE ADDRESS:		
	; ADDRESSEE: Howrey & Simon		
	; STREET: 1299 Pennsylvania Avenue N.W.		
	; CITY: Washington		
	; STATE: DC		
	; COUNTRY: USA		
	; ZIP: 20004		
	; COMPUTER READABLE FORM:		
	; MEDIUM TYPE: Diskette		
	; COMPUTER: IBM Compatible		
	; OPERATING SYSTEM: DOS		
	; SOFTWARE: FastSeq for Windows Version 2.0		
	; CURRENT APPLICATION DATA:		
	; APPLICATION NUMBER: US/09/037,751		
	; FILING DATE: 17-Dec-1999		
	; CLASSIFICATION: <Unknown>		
	; PRIOR APPLICATION DATA:		
	; APPLICATION NUMBER: 09/037,751		
	; FILING DATE: <Unknown>		
	; ATTORNEY/AGENT INFORMATION:		
	; NAME: Halluin, Albert P		
	; REGISTRATION NUMBER: 25,277		
	; REFERENCE/DOCKET NUMBER: 00801.0140.999		
	; TELECOMMUNICATION INFORMATION:		
	; TELEPHONE: 650-463-8109		
	; TELEFAX: 650-463-8400		
	; TELEX: <Unknown>		
	; INFORMATION FOR SEQ ID NO: 1:		
	SEQUENCE CHARACTERISTICS:		
	LENGTH: 6395 base pairs		
	TYPE: nucleic acid		
	STRANDEDNESS: single		
	TOPOLOGY: unknown		
	MOLECULE TYPE: Genomic RNA		
	SEQUENCE DESCRIPTION: SEQ ID NO: 1:		
	US-09-466-422-1		
	Query Match	55.8%; Score 42.4; DB 3; Length 63	





```

; TOPOLOGY: unknown
; MOLECULE TYPE: Genomic RNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-362-527-3

Query Match          55.8%; Score 42.4; DB 4; Length 6435;
Best Local Similarity 56.8%; Pred. No. 4.3e-06;
Matches 43; Conservative 12; Mismatches 21; Indels 0; Gaps 0;

Qy      1  GTGACAGACGGCTCGCCAAATGAACTCACCTGAAAAGGTTGTTGAGGAGTTCGTGGATGAA 60
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      5443  GUGAGAGACGGAGGGCCCAUGGAACUUAACAGAGAAGCUGCUGAUGAGUUAUGGAAGAU 5502

Qy      61  GTACCAATGGCTGTGA 76
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      5503  GUCCCUAUGUGCGAUCA 5518

Search completed: January 17, 2005, 16:36:12
Job time : 5.08894 secs

```



DB 5490 GTACCAATGGCTGTGA 5505

```
RESULT 2
US-10-211-079-21
; Sequence 21, Application US/10211079
; Publication No. US20030148315A1
; GENERAL INFORMATION:
; APPLICANT: Padgett, Hal S.
; APPLICANT: Vaewhongs, Andrew A.
; APPLICANT: Vojdani, Fakhrieh S.
; APPLICANT: Smith, Mark L.
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding CEL I
; FILE REFERENCE: P-LG 5381
; CURRENT APPLICATION NUMBER: US/10/211,079
; PRIOR FILING DATE: 2002-08-01
; PRIOR APPLICATION NUMBER: US 10/098,155
; PRIOR FILING DATE: 2002-03-14
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 769
; TYPE: DNA
; ORGANISM: tobamovirus
US-10-211-079-21

Query Match          91.6%; Score 69.6; DB 15; Length 769;
Best Local Similarity 94.7%; Pred. No. 1.4e-14;
Matches 72; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GTGACAGCGCTCGCCAAATGAACCTCACTGAAAAGTTGTTGAGGAGTTCTGATGAA 60
Db 541 GTACACAGCGCTCGCCAAATGAACCTCACTGAAAAGTTGTTGAGGAGTTCTGATGAA 600
Qy 61 GTACCAATGGCTGTGA 76
Db 601 GTACCAATGGCTGTGA 616

RESULT 3
US-10-356-708-20
; Sequence 20, Application US/10356708
; Publication No. US20030157682A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; TITLE OF INVENTION: MISMATCH ENDONUCLEASES AND METHODS OF USE THEREOF
; FILE REFERENCE: P-LG 10100
; CURRENT APPLICATION NUMBER: US/10/356,708
; CURRENT FILING DATE: 2003-01-31
; PRIOR APPLICATION NUMBER: 60/353,722
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 10/098,155
; PRIOR FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: 10/211,079
; PRIOR FILING DATE: 2002-08-01
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 769
; TYPE: DNA
; ORGANISM: Tobacco mosaic virus-U2
US-10-356-708-20

Query Match          91.6%; Score 69.6; DB 15; Length 769;
Best Local Similarity 94.7%; Pred. No. 1.4e-14;
Matches 72; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GTGACAGCGCTCGCCAAATGAACCTCACTGAAAAGTTGTTGAGGAGTTCTGATGAA 60
Db 541 GTACACAGCGCTCGCCAAATGAACCTCACTGAAAAGTTGTTGAGGAGTTCTGATGAA 600
Qy 61 GTACCAATGGCTGTGA 76
Db 601 GTACCAATGGCTGTGA 616

RESULT 4
US-10-280-913A-20
; Sequence 20, Application US/10280913A
; Publication No. US20040110130A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; TITLE OF INVENTION: POPULATION OF POLYNUCLEOTIDE SEQUENCE VARIANTS
; FILE REFERENCE: LSBC-PLG5392-CIP
; CURRENT APPLICATION NUMBER: US/10/280,913A
; CURRENT FILING DATE: 2002-10-25
; PRIOR APPLICATION NUMBER: 60/402,342
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 10/066,390
; PRIOR FILING DATE: 2002-02-01
; PRIOR APPLICATION NUMBER: 60/268,785
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: 60/266,386
; PRIOR FILING DATE: 2002-02-02
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 769
; TYPE: DNA
; ORGANISM: Tobacco mosaic virus-U2
US-10-280-913A-20

Query Match          91.6%; Score 69.6; DB 17; Length 769;
Best Local Similarity 94.7%; Pred. No. 1.4e-14;
Matches 72; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GTGACAGCGCTCGCCAAATGAACCTCACTGAAAAGTTGTTGAGGAGTTCTGATGAA 60
Db 541 GTACACAGCGCTCGCCAAATGAACCTCACTGAAAAGTTGTTGAGGAGTTCTGATGAA 600
Qy 61 GTACCAATGGCTGTGA 76
Db 601 GTACCAATGGCTGTGA 616

RESULT 5
US-10-684-134-20
; Sequence 20, Application US/10684134
; Publication No. US20040142433A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; TITLE OF INVENTION: POPULATION OF POLYNUCLEOTIDE SEQUENCE VARIANTS
; FILE REFERENCE: LSBC-PLG5392-CIP
; CURRENT APPLICATION NUMBER: US/10/684,134
; CURRENT FILING DATE: 2003-10-10
; PRIOR APPLICATION NUMBER: 60/402,342
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 10/066,390
; PRIOR FILING DATE: 2002-02-01
; PRIOR APPLICATION NUMBER: 60/268,785
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: 60/266,386
; PRIOR FILING DATE: 2002-02-02
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 769
; TYPE: DNA
; ORGANISM: Tobacco mosaic virus-U2
US-10-684-134-20

Query Match          91.6%; Score 69.6; DB 17; Length 769;
Best Local Similarity 94.7%; Pred. No. 1.4e-14;
Matches 72; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GTGACAGCGCTCGCCAAATGAACCTCACTGAAAAGTTGTTGAGGAGTTCTGATGAA 60
Db 541 GTACACAGCGCTCGCCAAATGAACCTCACTGAAAAGTTGTTGAGGAGTTCTGATGAA 600
Qy 61 GTACCAATGGCTGTGA 76
Db 601 GTACCAATGGCTGTGA 616
```

```
Db 601 GTACCAATGGCTGTGA 616

RESULT 4
US-10-280-913A-20
; Sequence 20, Application US/10280913A
; Publication No. US20040110130A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; TITLE OF INVENTION: POPULATION OF POLYNUCLEOTIDE SEQUENCE VARIANTS
; FILE REFERENCE: LSBC-PLG5392-CIP
; CURRENT APPLICATION NUMBER: US/10/280,913A
; CURRENT FILING DATE: 2002-10-25
; PRIOR APPLICATION NUMBER: 60/402,342
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 10/066,390
; PRIOR FILING DATE: 2002-02-01
; PRIOR APPLICATION NUMBER: 60/268,785
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: 60/266,386
; PRIOR FILING DATE: 2002-02-02
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 769
; TYPE: DNA
; ORGANISM: Tobacco mosaic virus-U2
US-10-280-913A-20

Query Match          91.6%; Score 69.6; DB 17; Length 769;
Best Local Similarity 94.7%; Pred. No. 1.4e-14;
Matches 72; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GTGACAGCGCTCGCCAAATGAACCTCACTGAAAAGTTGTTGAGGAGTTCTGATGAA 60
Db 541 GTACACAGCGCTCGCCAAATGAACCTCACTGAAAAGTTGTTGAGGAGTTCTGATGAA 600
Qy 61 GTACCAATGGCTGTGA 76
Db 601 GTACCAATGGCTGTGA 616

RESULT 5
US-10-684-134-20
; Sequence 20, Application US/10684134
; Publication No. US20040142433A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; TITLE OF INVENTION: POPULATION OF POLYNUCLEOTIDE SEQUENCE VARIANTS
; FILE REFERENCE: LSBC-PLG5392-CIP
; CURRENT APPLICATION NUMBER: US/10/684,134
; CURRENT FILING DATE: 2003-10-10
; PRIOR APPLICATION NUMBER: 60/402,342
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 10/066,390
; PRIOR FILING DATE: 2002-02-01
; PRIOR APPLICATION NUMBER: 60/268,785
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: 60/266,386
; PRIOR FILING DATE: 2002-02-02
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 769
; TYPE: DNA
; ORGANISM: Tobacco mosaic virus-U2
US-10-684-134-20

Query Match          91.6%; Score 69.6; DB 17; Length 769;
Best Local Similarity 94.7%; Pred. No. 1.4e-14;
Matches 72; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GTGACAGCGCTCGCCAAATGAACCTCACTGAAAAGTTGTTGAGGAGTTCTGATGAA 60
Db 541 GTACACAGCGCTCGCCAAATGAACCTCACTGAAAAGTTGTTGAGGAGTTCTGATGAA 600
Qy 61 GTACCAATGGCTGTGA 76
Db 601 GTACCAATGGCTGTGA 616
```



Db 541 GTACACAGCGCTCGCCCAATTGAACTCACTGAAAAGTTGTTGAGGAGTTCATAGATGAA 600  
Qy 61 GTACCAATGGCTGTGA 76  
Db 601 GTACCAATGGCTGTGA 616

## RESULT 6

US-10-637-758-20  
; Sequence 20, Application US/10637758  
; Publication No. US20040180352A1  
; GENERAL INFORMATION:  
; APPLICANT: Large Scale Biology Corporation  
; TITLE OF INVENTION: POPULATION OF POLYNUCLEOTIDE SEQUENCE VARIANTS  
; FILE REFERENCE: LSBC-PLG5392-CIP  
; CURRENT APPLICATION NUMBER: US/10/637,758  
; PRIOR FILING DATE: 2003-08-08  
; PRIOR APPLICATION NUMBER: US/10/280,913  
; PRIOR FILING DATE: 2002-10-25  
; PRIOR APPLICATION NUMBER: 60/402,342  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 10/066,390  
; PRIOR FILING DATE: 2002-02-01  
; PRIOR APPLICATION NUMBER: 60/268,785  
; PRIOR FILING DATE: 2001-02-14  
; PRIOR APPLICATION NUMBER: 60/266,386  
; PRIOR FILING DATE: 2002-02-02  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 20  
; LENGTH: 769  
; TYPE: DNA  
; ORGANISM: Tobacco mosaic virus-U2  
US-10-637-758-20

Query Match 91.6%; Score 69.6; DB 17; Length 769;  
Best Local Similarity 94.7%; Pred. No. 1.4e-14;  
Matches 72; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GTGACAGACGCGCTCGCCCAATTGAACTCACTGAAAAGTTGTTGAGGAGTTCGTGGATGAA 60  
Db 541 GTACACAGCGCTCGCCCAATTGAACTCACTGAAAAGTTGTTGAGGAGTTCATAGATGAA 600  
Qy 61 GTACCAATGGCTGTGA 76  
Db 601 GTACCAATGGCTGTGA 616

## RESULT 7

US-10-211-079-27/c  
; Sequence 27, Application US/10211079  
; Publication No. US20030148315A1  
; GENERAL INFORMATION:  
; APPLICANT: Padgett, Hal S.  
; APPLICANT: Vaewhongs, Andrew A.  
; APPLICANT: Vojdani, Fakhrish S.  
; APPLICANT: Smith, Mark L.  
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding CEL I  
; FILE REFERENCE: P-LG 5381  
; CURRENT APPLICATION NUMBER: US/10/211,079  
; PRIOR FILING DATE: 2002-08-01  
; PRIOR APPLICATION NUMBER: US 10/098,155  
; PRIOR FILING DATE: 2002-03-14  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 27  
; LENGTH: 772  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: synthetic construct

## US-10-211-079-27

Query Match 89.5%; Score 68; DB 15; Length 772;  
Best Local Similarity 93.4%; Pred. No. 5.2e-14;  
Matches 71; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
Qy 1 GTGACAGACGCGCTCGCCCAATTGAACTCACTGAAAAGTTGTTGAGGAGTTCGTGGATGAA 60  
Db 232 GTACACAGCGCTCGCCCAATTGAACTCACTGAAAAGTTGTTGATGATGATAGATGAA 173  
Qy 61 GTACCAATGGCTGTGA 76  
Db 172 GTACCAATGGCTGTGA 157

## RESULT 8

US-10-356-708-26/c  
; Sequence 26, Application US/10356708  
; Publication No. US20030157682A1  
; GENERAL INFORMATION:  
; APPLICANT: Large Scale Biology Corporation  
; TITLE OF INVENTION: MISMATCH ENDONUCLEASES AND METHODS OF USE THEREOF  
; FILE REFERENCE: P-LG 10100  
; CURRENT APPLICATION NUMBER: US/10/356,708  
; CURRENT FILING DATE: 2003-01-31  
; PRIOR APPLICATION NUMBER: 60/353,722  
; PRIOR FILING DATE: 2000-02-01  
; PRIOR APPLICATION NUMBER: 10/098,155  
; PRIOR FILING DATE: 2002-03-14  
; PRIOR APPLICATION NUMBER: 10/211,079  
; PRIOR FILING DATE: 2002-08-01  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 26  
; LENGTH: 772  
; TYPE: DNA  
; ORGANISM: Artificial Sequence derived from TMV & TMV-U2  
; FEATURE:  
; OTHER INFORMATION: This sequence was derived by shuffling in accordance with the method of the present invention.  
US-10-356-708-26

Query Match 89.5%; Score 68; DB 15; Length 772;  
Best Local Similarity 93.4%; Pred. No. 5.2e-14;  
Matches 71; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
Qy 1 GTGACAGACGCGCTCGCCCAATTGAACTCACTGAAAAGTTGTTGAGGAGTTCGTGGATGAA 60  
Db 232 GTACACAGCGCTCGCCCAATTGAACTCACTGAAAAGTTGTTGATGATGATAGATGAA 173  
Qy 61 GTACCAATGGCTGTGA 76  
Db 172 GTACCAATGGCTGTGA 157

## RESULT 9

US-10-280-913A-26/c  
; Sequence 26, Application US/10280913A  
; Publication No. US20040110130A1  
; GENERAL INFORMATION:  
; APPLICANT: Large Scale Biology Corporation  
; TITLE OF INVENTION: POPULATION OF POLYNUCLEOTIDE SEQUENCE VARIANTS  
; FILE REFERENCE: LSBC-PLG5392-CIP  
; CURRENT APPLICATION NUMBER: US/10/280,913A  
; CURRENT FILING DATE: 2002-10-25  
; PRIOR APPLICATION NUMBER: 60/402,342  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 10/066,390  
; PRIOR FILING DATE: 2002-02-01  
; PRIOR APPLICATION NUMBER: 60/268,785  
; PRIOR FILING DATE: 2001-02-14  
; PRIOR APPLICATION NUMBER: 60/266,386  
; PRIOR FILING DATE: 2002-02-02

NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 26  
; LENGTH: 772  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: This construct was derived by GRAMMR shuffling in accordance with  
; OTHER INFORMATION: the methodology of the present invention.  
US-10-280-913A-26

Query Match 89.5%; Score 68; DB 17; Length 772;  
Best Local Similarity 93.4%; Pred. No. 5.2e-14;  
Matches 71; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
Qy 1 GTGACAGACGGCTCGCCCAATTGAACCTCACTGAAAGGTTGTTGAGGAGTTGCTGGATGAA 60  
Db 232 GTACACAGCGCTCGCCCAATTGAACCTCACTGAAAGGTTGTTGAGGAGTTGCTGGATGAA 173  
Qy 61 GTACCAATGGCTGTGA 76  
Db 172 GTACCAATGGCTGTGA 157

RESULT 10  
US-10-684-134-26/c  
; Sequence 26, Application US/10684134  
; Publication No. US20040142433A1  
; GENERAL INFORMATION:  
; APPLICANT: Large Scale Biology Corporation  
; TITLE OF INVENTION: POPULATION OF POLYNUCLEOTIDE SEQUENCE VARIANTS  
; FILE REFERENCE: LSBC-PLG5392-CIP  
; CURRENT APPLICATION NUMBER: US/10/684.134  
; CURRENT FILING DATE: 2003-10-10  
; PRIOR APPLICATION NUMBER: 60/402,342  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 10/066,390  
; PRIOR FILING DATE: 2002-02-01  
; PRIOR APPLICATION NUMBER: 60/268,785  
; PRIOR FILING DATE: 2001-02-14  
; PRIOR APPLICATION NUMBER: 60/266,386  
; PRIOR FILING DATE: 2002-02-02  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 26  
; LENGTH: 772  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: This construct was derived by GRAMMR shuffling in accordance with  
; OTHER INFORMATION: the methodology of the present invention.  
US-10-684-134-26

Query Match 89.5%; Score 68; DB 17; Length 772;  
Best Local Similarity 93.4%; Pred. No. 5.2e-14;  
Matches 71; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
Qy 1 GTGACAGACGGCTCGCCCAATTGAACCTCACTGAAAGGTTGTTGAGGAGTTGCTGGATGAA 60  
Db 232 GTACACAGCGCTCGCCCAATTGAACCTCACTGAAAGGTTGTTGAGGAGTTGCTGGATGAA 173  
Qy 61 GTACCAATGGCTGTGA 76  
Db 172 GTACCAATGGCTGTGA 157

RESULT 11  
US-10-637-758-26/c  
; Sequence 26, Application US/10637758  
; Publication No. US20040180352A1  
; GENERAL INFORMATION:  
; APPLICANT: Large Scale Biology Corporation  
; TITLE OF INVENTION: POPULATION OF POLYNUCLEOTIDE SEQUENCE VARIANTS

FILE REFERENCE: LSBC-PLG5392-CIP  
; CURRENT APPLICATION NUMBER: US/10/637,758  
; CURRENT FILING DATE: 2003-08-08  
; PRIOR APPLICATION NUMBER: US/10/280,913  
; PRIOR FILING DATE: 2002-10-25  
; PRIOR APPLICATION NUMBER: 60/402,342  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 10/066,390  
; PRIOR FILING DATE: 2002-02-01  
; PRIOR APPLICATION NUMBER: 60/268,785  
; PRIOR FILING DATE: 2001-02-14  
; PRIOR APPLICATION NUMBER: 60/266,386  
; PRIOR FILING DATE: 2002-02-02  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 26  
; LENGTH: 772  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: This construct was derived by GRAMMR shuffling in accordance with  
; OTHER INFORMATION: the methodology of the present invention.  
US-10-637-758-26

Query Match 89.5%; Score 68; DB 17; Length 772;  
Best Local Similarity 93.4%; Pred. No. 5.2e-14;  
Matches 71; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
Qy 1 GTGACAGACGGCTCGCCCAATTGAACCTCACTGAAAGGTTGTTGAGGAGTTGCTGGATGAA 60  
Db 232 GTACACAGCGCTCGCCCAATTGAACCTCACTGAAAGGTTGTTGAGGAGTTGCTGGATGAA 173  
Qy 61 GTACCAATGGCTGTGA 76  
Db 172 GTACCAATGGCTGTGA 157

RESULT 12  
US-10-211-079-26/c  
; Sequence 26, Application US/10211079  
; Publication No. US20030148315A1  
; GENERAL INFORMATION:  
; APPLICANT: Padgett, Hal S.  
; APPLICANT: Vaewhongs, Andrew A.  
; APPLICANT: Vojdani, Fakhrhien S.  
; APPLICANT: Smith, Mark L.  
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding CEL I  
; FILE REFERENCE: P-LG 5381  
; CURRENT APPLICATION NUMBER: US/10/211,079  
; CURRENT FILING DATE: 2002-08-01  
; PRIOR APPLICATION NUMBER: US 10/098,155  
; PRIOR FILING DATE: 2002-03-14  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 26  
; LENGTH: 769  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: synthetic construct  
US-10-211-079-26

Query Match 64.2%; Score 48.8; DB 15; Length 769;  
Best Local Similarity 77.6%; Pred. No. 3.6e-07;  
Matches 59; Conservative 0; Mismatches 17; Indels 0; Gaps 0;  
Qy 1 GTGACAGACGGCTCGCCCAATTGAACCTCACTGAAAGGTTGTTGAGGAGTTGCTGGATGAA 60  
Db 229 GTGAACGATGGAGGACCCCATGAACTTTCAGAAAGTTGTTGATGAGTTTCATAGATGAA 170  
Qy 61 GTACCAATGGCTGTGA 76

```
Db      169 GTACCAATGGCTGTGA 154

RESULT 13
US-10-356-708-25/c
; Sequence 25, Application US/10356708
; Publication No. US20030157682A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; TITLE OF INVENTION: MISMATCH ENDONUCLEASES AND METHODS OF USE THEREOF
; FILE REFERENCE: P-LG 10100
; CURRENT APPLICATION NUMBER: US/10/356,708
; CURRENT FILING DATE: 2003-01-31
; PRIOR FILING DATE: 2003-01-31
; PRIOR FILING DATE: 2000-02-01
; PRIOR FILING DATE: 2000-02-01
; PRIOR FILING DATE: 2002-03-14
; PRIOR FILING DATE: 2002-03-14
; PRIOR FILING DATE: 2002-08-01
; PRIOR FILING DATE: 2002-08-01
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25
; LENGTH: 769
; TYPE: DNA
; ORGANISM: Artificial Sequence derived from ToMV & TMV-U2
; FEATURE:
; OTHER INFORMATION: This sequence was derived by shuffling in accordance with the method
; OTHER INFORMATION: hology of the present invention.
US-10-356-708-25

Query Match      64.2%; Score 48.8; DB 15; Length 769;
Best Local Similarity 77.6%; Pred. No. 3.6e-07;
Matches 59; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy      1 GTGACAGACGGCTCGCCAATTGAACTCACTCACTGAAAGGTTGTTGAGGAGTTTCGTGGATGAA 60
        |||||
Db      229 GTGACAGATGGAGACCCATGGAACCTTCAGAGAAGTTGTTGATGAGTTTCATAGATGAA 170

Qy      61 GTACCAATGGCTGTGA 76
        |||||
Db      169 GTACCAATGGCTGTGA 154

RESULT 14
US-10-280-913A-25/c
; Sequence 25, Application US/10280913A
; Publication No. US20040110130A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; TITLE OF INVENTION: POPULATION OF POLYNUCLEOTIDE SEQUENCE VARIANTS
; FILE REFERENCE: LSBC-PLG5392-CIP
; CURRENT APPLICATION NUMBER: US/10/280,913A
; CURRENT FILING DATE: 2002-10-25
; PRIOR APPLICATION NUMBER: 60/402,342
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 10/066,390
; PRIOR FILING DATE: 2002-02-01
; PRIOR APPLICATION NUMBER: 60/268,785
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: 60/266,386
; PRIOR FILING DATE: 2002-02-02
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25
; LENGTH: 769
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: This construct was derived by GRAMMR shuffling in accordance with
; OTHER INFORMATION: the methodology of the present invention.
US-10-280-913A-25

Query Match      64.2%; Score 48.8; DB 17; Length 769;
Best Local Similarity 77.6%; Pred. No. 3.6e-07;
Matches 59; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy      1 GTGACAGACGGCTCGCCAATTGAACTCACTCACTGAAAGGTTGTTGAGGAGTTTCGTGGATGAA 60
        |||||
Db      229 GTGACAGATGGAGACCCATGGAACCTTCAGAGAAGTTGTTGATGAGTTTCATAGATGAA 170

Qy      61 GTACCAATGGCTGTGA 76
        |||||
Db      169 GTACCAATGGCTGTGA 154

RESULT 15
US-10-684-134-25/c
; Sequence 25, Application US/10684134
; Publication No. US20040142433A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; TITLE OF INVENTION: POPULATION OF POLYNUCLEOTIDE SEQUENCE VARIANTS
; FILE REFERENCE: LSBC-PLG5392-CIP
; CURRENT APPLICATION NUMBER: US/10/684,134
; CURRENT FILING DATE: 2003-10-10
; PRIOR APPLICATION NUMBER: 60/402,342
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 10/066,390
; PRIOR FILING DATE: 2002-02-01
; PRIOR APPLICATION NUMBER: 60/268,785
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: 60/266,386
; PRIOR FILING DATE: 2002-02-02
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25
; LENGTH: 769
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: This construct was derived by GRAMMR shuffling in accordance with
; OTHER INFORMATION: the methodology of the present invention.
US-10-684-134-25

Query Match      64.2%; Score 48.8; DB 17; Length 769;
Best Local Similarity 77.6%; Pred. No. 3.6e-07;
Matches 59; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy      1 GTGACAGACGGCTCGCCAATTGAACTCACTCACTGAAAGGTTGTTGAGGAGTTTCGTGGATGAA 60
        |||||
Db      229 GTGACAGATGGAGACCCATGGAACCTTCAGAGAAGTTGTTGATGAGTTTCATAGATGAA 170

Qy      61 GTACCAATGGCTGTGA 76
        |||||
Db      169 GTACCAATGGCTGTGA 154

Search completed: January 17, 2005, 21:46:50
Job time : 26.3891 secs
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 17, 2005, 05:56:43 ; Search time 150.251 Seconds  
(without alignments)  
18431.958 Million cell updates/sec

Title: US-09-551-494-5\_COPY\_5430\_5505

Perfect score: 76  
Sequence: 1 gtgacagagcgtcgcaat.....tgaagtacaaatggctgtga 76

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_hic:\*  
4: gb\_est3:\*  
5: gb\_est4:\*  
6: gb\_est5:\*  
7: gb\_est6:\*  
8: gb\_gest1:\*  
9: gb\_gest2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	39.4	51.8	188	4	BM068137 KS08017C0
c 2	39.4	51.8	348	4	BM067518 KS08006E1
c 3	32.6	42.9	583	6	CB264749 41-E01466
4	31.6	41.6	954	8	B12288 T2M2-Sp6 TA
5	31	40.8	389	1	AV441961 AV441961
6	31	40.8	530	6	CA963974 CATI008D1
7	31	40.8	591	9	FR0020421 F.rubripe
8	31	40.8	656	5	B0636714 010F08 In
9	31	40.8	686	5	B0636451 052H05 In
10	31	40.8	1535	3	CNS0ADUO Arabidops
11	31	40.8	1579	3	CNS0ACVY Arabidops
c 12	30.2	39.7	711	5	EX845438 BX845438
c 13	30.2	39.7	800	5	B0915084 AGENCOURT
c 14	30	39.5	446	9	AL013342 F.rubripe
c 15	29.8	39.2	561	5	BX553578 BX553578
c 16	29.6	38.9	357	9	CG742490 ZMMBBc019
17	29.6	38.9	854	7	CO800453 AGENCOURT
c 18	29.4	38.7	732	1	AU305571 AU305571
c 19	28.8	37.9	387	5	B0816956 UAL1BPB12
c 20	28.8	37.9	398	9	CNS000STV Arabidops
c 21	28.8	37.9	460	7	CK095490 UAL1BPB12
c 22	28.8	37.9	498	8	AZ912881 RPCI-24-1
c 23	28.8	37.9	596	1	AJ769214 AJ769214
24	28.8	37.9	605	1	AJ774675 AJ774675

25	28.8	37.9	643	1	AJ767586
c 26	28.8	37.9	946	8	CC099545
c 27	28.6	37.6	514	8	BZ590961
c 28	28.6	37.6	516	8	BZ590922
c 29	28.6	37.6	522	8	BZ590931
c 30	28.6	37.6	527	8	BZ588716
c 31	28.6	37.6	529	8	BZ586702
c 32	28.6	37.6	548	7	CN947691
c 33	28.6	37.6	534	8	BZ586917
c 34	28.6	37.6	763	9	CC621362
c 35	28.6	37.6	877	9	CG305892
c 36	28.6	37.6	1003	8	BZ700933
c 37	28.2	37.1	385	9	CG742799
38	28.2	37.1	616	9	CC681582
39	28.2	37.1	635	9	CG392856
40	28.2	37.1	938	9	CL238374
41	28	36.8	244	9	CL226782
42	28	36.8	264	9	CG838137
43	28	36.8	276	9	CG743187
44	28	36.8	300	9	CG831760
45	28	36.8	547	6	CD836299

#### ALIGNMENTS

RESULT 1  
BM068137/c  
LOCUS BM068137 188 bp mRNA linear EST 11-SEP-2002  
DEFINITION KS08017C07 KS08 Capsicum annum cdNA, mRNA sequence.  
ACCESSION BM068137  
VERSION BM068137.1 GI:22788242  
KEYWORDS EST.  
SOURCE Capsicum annum  
ORGANISM Capsicum annum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
asterids; lamids; Solanales; Solanaceae; Capsicum.  
REFERENCE 1 (bases 1 to 188)  
AUTHORS Lee, S., Kim, S.-Y., Chung, Y.-H., Shin, H.-J., Goh, S.-H., Pai, H.-S.,  
Hur, C.-G. and Choi, D.  
TITLE Generation of Expressed Sequence Tags from Hot Pepper (Capsicum  
annuum L.) and Sequence Analysis in Relation to Hypersensitive  
Response Against Pathogen  
JOURNAL Unpublished (2001)  
COMMENT Contact: Doil Choi  
Genome Research Institute of Bioscience and Biotechnology  
P.O. Box 115, Yusong, Taejeon, 305-600, Republic of Korea  
Tel: 82-42-860-4340  
Fax: 82-42-860-4309  
Email: doil@mail.kribb.re.kr  
High quality sequence stop: 188.

FEATURES  
source  
1..188  
/organism="Capsicum annum"  
/mol\_type="mRNA"  
/cultivar="Hang Keun"  
/db\_xref="taxon:4072"  
/tissue type="anther"  
/dev\_stage="10 weeks after germination"  
/clone\_lib="KS08"  
/note="Vector: pBluescript SK(-)"

ORIGIN  
Query Match 51.8%; Score 39.4; DB 4; Length 188;  
Best Local Similarity 71.2%; Pred.No. 0.01;  
Matches 52; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Qy 1 GTGACAGACGGCTCGCCAAATGAACCTCACTGAAAGGTTGTTGAGGAGTTGTTGAGGAGTCA 60  
|||||  
Db 122 GTGTGGAGGAGGAGGCCCGTTGAACCTTACAGAGCACTGTTGATGATTCATCGAATCA 63

QY 61 GTACCAATGGCTG 73  
Db 62 GTTCCAATGGCTG 50

RESULT 2  
BM067518/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

BM067518 348 bp mRNA linear EST 11-SEP-2002  
KS08006E10 KS08 Capsicum annuum cDNA, mRNA sequence.  
BM067518  
EST.  
BM067518.1 GI:22787638  
Capsicum annuum  
Capsicum annuum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
asterids; lamids; Solanales; Solanaceae; Capsicum.  
1 (bases 1 to 348)  
Lee, S.-G., Kim, S.-Y., Chung, Y.-H., Shin, H.-J., Goh, S.-H., Pai, H.-S.,  
Hur, C.-G. and Choi, D.  
Generation of Expressed Sequence Tags from Hot Pepper (Capsicum  
annuum L.) and Sequence Analysis in Relation to Hypersensitive  
Response Against Pathogen  
Unpublished (2001)  
Contact: Doil Choi  
Genome Research Center and National Center for Genome Information  
Korea Research Institute of Bioscience and Biotechnology  
P.O. Box 115, Yusong, Taejeon, 305-600, Republic of Korea  
Tel: 82-42-860-4340  
Fax: 82-42-860-4309  
Email: doil@mail.kribb.re.kr  
High quality sequence stop: 348.  
Location/Qualifiers  
1..348  
/organism="Capsicum annuum"  
/mol\_type="mRNA"  
/cultivar="Hang Keun"  
/db\_xref="taxon:4072"  
/tissue\_type="anther"  
/dev\_stage="10 weeks after germination"  
/clone\_lib="KS08"  
/note="Vector: pBluescript SK(-)"

FEATURES  
source  
1..348  
/organism="Capsicum annuum"  
/mol\_type="mRNA"  
/cultivar="Hang Keun"  
/db\_xref="taxon:4072"  
/tissue\_type="anther"  
/dev\_stage="10 weeks after germination"  
/clone\_lib="KS08"  
/note="Vector: pBluescript SK(-)"

ORIGIN  
Query Match 51.8%; Score 39.4; DB 4; Length 348;  
Best Local Similarity 71.2%; Pred. No. 0.011;  
Matches 52; Conservative 0; Mismatches 21; Indels 0; Gaps 0;  
QY 1 GTGACAGCGCTCGCCAAATTGAACCTCACTGAAAGGTTGTTGAGGAGTTGCGTGATGAA 60  
Db 294 GTGTGGAGAGGAGGACCGCTTGAACCTTACAGACGAGTTGTTGATGAGTTTCATCGAATCA 235  
QY 61 GTACCAATGGCTG 73  
Db 234 GTTCCAATGGCTG 222

RESULT 3  
CB264749  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

CB264749 583 bp mRNA linear EST 06-NOV-2003  
41-E014660-035-002-B11-T7R MP1Z-ADIS-035 Arabidopsis thaliana cDNA  
clone MP1Zp2000B11Q 5-PRIME, mRNA sequence.  
CB264749  
EST.  
CB264749.1 GI:32889522  
Arabidopsis thaliana (thale cress)  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
1 (bases 1 to 583)  
Schmid, K.J., Soerensen, T.R., Stracke, R., Torjek, O., Altmann, T.,  
Mitchell-Olds, T. and Weisshaar, B.

TITLE  
JOURNAL  
MEDLINE  
PUBMED  
COMMENT

Large-scale identification and analysis of genome-wide  
single-nucleotide polymorphisms for mapping in Arabidopsis thaliana  
Genome Res. 13 (6), 1250-1257 (2003)  
22683290  
12799357  
Contact: Weisshaar B  
ADIS DNA core facility at MP1Z  
Max-Planck-Institute for Plant Breeding Research  
Carl-von-Linne Weg 10, 50829 Koeln, Germany  
Fax: 00492215062851  
Email: weisshaar@mpiz-koeln.mpg.de  
Insert Length: 583 Std Error: 0.00  
Plate: 2 row: B column: 11  
Seq primer: T7R; CTAATACGACTCACTATAGGA.  
Location/Qualifiers  
1..583  
/organism="Arabidopsis thaliana"  
/mol\_type="mRNA"  
/cultivar="Achkarren-2 (AK-2)"  
/db\_xref="GABI:594666"  
/db\_xref="taxon:3702"  
/clone="MP1Zp2000B11Q"  
/tissue\_type="inflorescence"  
/lab\_host="E. coli TOP10"  
/clone\_lib="MP1Z-ADIS-035"  
/note="Vector: pSPORT1; Site 1: SalI; Site 2: NotI; cDNA  
library from Arabidopsis thaliana, accession Achkarren-2;  
inflorescences from flower buds to young siliques; library  
was made at the Max-Planck-Institute for Plant Breeding  
Research, Cologne, Germany; cloning sites SalI-NotI,  
primer sites and orientation:  
T7-SalI-CCACGCTCCG-5prime-cDNA-polyA-CC-NotI-SP6; Note:  
Sequencing granted in the context of the GABI Arabidopsis  
Verbund I: Genetic Diversity, 'Establishment of  
high-efficiency SNP-based mapping tools and development of  
methods for genome-wide mutation detection' PI: Bernd  
Weisshaar; Sequence submission managed by  
RZPD/GABI-Primary database: http://gabi.rzpd.de This clone  
is available from RZPD; contact RZPD (clone@rzpd.de) for  
further information."

ORIGIN  
Query Match 42.9%; Score 32.6; DB 6; Length 583;  
Best Local Similarity 66.2%; Pred. No. 2.6;  
Matches 47; Conservative 0; Mismatches 24; Indels 0; Gaps 0;  
QY 6 AGACGCTCGCCAAATTGAACCTCACTGAAAGGTTGTTGAGGAGTTGCGTGATGATGATCC 65  
Db 381 AGATGCTTCTCCACTTGAATCATGATGAAGCTCTTGAGAAATTTGGAGACGAAATCGC 440  
QY 66 AATGGCTGTGA 76  
Db 441 AATTGCTTCA 451

RESULT 4  
B12288  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

B12288 954 bp DNA linear GSS 14-MAY-1997  
T2M2-Sp6 TAMU Arabidopsis thaliana genomic clone T2M2, genomic  
survey sequence.  
B12288  
B12288.1 GI:2093409  
GSS.  
Arabidopsis thaliana (thale cress)  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
1 (bases 1 to 954)  
Feng, J., Dewar, K., Buehler, E., Kim, C., Li, Y., Shinn, P., Sun, H. and  
Ecker, J.  
BAC End Sequences at ATGC  
Unpublished (1997)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

```

COMMENT
Other GSSs: T2M2-T7
Contact: Ecker J.
Arabidopsis thaliana Genome Center
University of Pennsylvania
Dept. of Biology, University of Pennsylvania, Philadelphia, PA
19104
Tel: 215-898-9384
Fax: 215-898-8780
Email: jecker@genome.bio.upenn.edu
Seq primer: Sp6
Class: BAC ends
High quality sequence start: 93
High quality sequence stop: 103.
FEATURES
    source
        1..934
            /organism="Arabidopsis thaliana"
            /mol_type="genomic DNA"
            /ecotype="Columbia"
            /db_xref="taxon:3702"
            /clone="T2M2"
            /sex="hermaphrodite"
            /clone_lib="TAMU"
            /notes="Vector: BelobACII; Site 1: HindIII; Site 2:
            HindIII; Produced by Rod Wing"
ORIGIN
Query Match 41.6%; Score 31.6; DB 8; Length 954;
Best Local Similarity 64.8%; Pred. No. 6.3;
Matches 46; Conservative 0; Mismatches 25; Indels 0; Gaps 0;
Qy 6 AGACGGCTCGCCCAATGAACCTCACTGAAAGGTTGTGAGGAGTTCGTGGATGAAGTACC 65
Db 72 ANAATGCTCTCCACTTGAATCATGATAAAGCTCTTGAGAGATTCGGAGAACAAATTCGC 131
Qy 66 AATGGCTGTGA 76
Db 132 AATGCTTTTA 142
RESULT 5
AV441961
LOCUS
DEFINITION
AV441961 Arabidopsis thaliana above-ground organ two to six-week
old Arabidopsis thaliana cDNA clone APD19f09_2 5', mRNA sequence.
ACCESSION
AV441961
VERSION
AV441961.1 GI:7612359
KEYWORDS
EST.
SOURCE
Arabidopsis thaliana (thale cress)
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
1 (bases 1 to 389)
Asamizu E., Nakamura Y., Sato S. and Tabata S.
A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries
DNA Res. 7 (3), 175-180 (2000)
JOURNAL
MEDLINE
20363093
PUBMED
10907847
COMMENT
Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
FEATURES
    source
        1..389
            /organism="Arabidopsis thaliana"
            /mol_type="mRNA"
            /ecotype="Columbia"
            /db_xref="taxon:3702"
            /clone="APD19f09_r"
            /tissue_type="aboveground organs"

```

---

```

/dev_stage="two to six-week old"
/clone_lib="Arabidopsis thaliana above-ground organ two to
six-week old"
/notes="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"
ORIGIN
Query Match 40.8%; Score 31; DB 1; Length 389;
Best Local Similarity 64.8%; Pred. No. 8.5;
Matches 46; Conservative 0; Mismatches 25; Indels 0; Gaps 0;
Qy 6 AGACGGCTCGCCCAATGAACCTCACTGAAAGGTTGTGAGGAGTTCGTGGATGAAGTACC 65
Db 62 AGATGCTTCTCCACTTGAATCATGATAAAGCTCTTGAGAGATTCGGAGAACAAATCGC 121
Qy 66 AATGGCTGTGA 76
Db 122 AATGCTTTTA 132
RESULT 6
CA963974
LOCUS
DEFINITION
CA963974 530 bp mRNA linear EST 03-JAN-2003
CATION08D10AF Infected Arabidopsis Leaf Arabidopsis thaliana cDNA,
mRNA sequence.
ACCESSION
CA963974
VERSION
CA963974.1 GI:27490531
KEYWORDS
EST.
SOURCE
Arabidopsis thaliana (thale cress)
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
1 (bases 1 to 530)
Lundgaard M., Emmersen J., Nielsen K.L., Wilson I., Somerville S.
and Welinder K.G.
EST sequencing of Erysiphe cichoracearum infected Arabidopsis
plants
Unpublished (2002)
JOURNAL
COMMENT
Contact: Karen G. Welinder
Institut for bioteknologi
Aalborg Universitet
Schlegelholmsvej 49, 9000 Aalborg, Denmark
Tel: +45 96358467
Fax: +45 98141808
Email: kgw@bio.auc.dk.
FEATURES
    source
        1..530
            /organism="Arabidopsis thaliana"
            /mol_type="mRNA"
            /ecotype="Columbia"
            /db_xref="taxon:3702"
            /dev_stage="Plant 3 weeks old, three days post infection"
            /clone_lib="Infected Arabidopsis Leaf"
            /notes="Organ: Leaf; Vector: pBluescript; Mixed cDNA
            library of Arabidopsis and E. cichoracearum infected leaf
            from three weeks old Arabidopsis plants. Plants were
            harvested 3 days after infection and mRNA oligo dt
            selected."
ORIGIN
Query Match 40.8%; Score 31; DB 6; Length 530;
Best Local Similarity 64.8%; Pred. No. 9;
Matches 46; Conservative 0; Mismatches 25; Indels 0; Gaps 0;
Qy 6 AGACGGCTCGCCCAATGAACCTCACTGAAAGGTTGTGAGGAGTTCGTGGATGAAGTACC 65
Db 215 AGATGCTTCTCCACTTGAATCATGATAAAGCTCTTGAGAGATTCGGAGAACAAATCGC 274
Qy 66 AATGGCTGTGA 76
Db 275 AATGCTTTTA 285

```

```

RESULT 7
FR0020421
LOCUS      F.rubripes GSS sequence, clone 041P11dB11, genomic survey sequence.
ACCESSION AL013304
VERSION    AL013304.1 GI:2679672
KEYWORDS   GSS; Genome survey sequence.
SOURCE     Takifugu rubripes (Fugu rubripes)
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodontidae; Takifugu.
REFERENCE
AUTHORS    Elgar, G., Clark, M.S., Meek, S., Smith, S., Warner, S., Williams, G. and
            Bouchireb, N., Cottage, A., Yeo, G.S., Umrana, Y., Williams, G. and
            Brenner, S.
TITLE      Generation and analysis of 25 Mb of genomic DNA from the pufferfish
JOURNAL    Fugu rubripes by sequence scanning
MEDLINE    Genome Res. 9 (10), 960-971 (1999)
PUBMED     99455097
REFERENCE  10523524
AUTHORS    Elgar, G., Clark, M., Smith, S., Meek, S., Warner, S., Umrana, Y.,
            Williams, G. and Brenner, S.
TITLE      Direct Submission
JOURNAL    Submitted (08-DEC-1997) MRC Human Genome Mapping Project Resource
COMMENT    Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@gmp.mrc.ac.uk
            Vector: pBluescript II KS
            V type: phagemid
            PRIMER: KS
DESCR      One pass dye-terminator sequencing of cosmid cloned genomic
            sequence.
FEATURES   source
            1..591
            /location=Qualifiers
            /mol_type="Takifugu rubripes"
            /organism="Takifugu rubripes"
            /db_type="genomic DNA"
            /db_xref="taxon:31033"
            /clones="041P11dB11"
            /clone_lib="cosmid 041P11"
ORIGIN
Query Match      40.8%; Score 31; DB 9; Length 591;
Best Local Similarity 64.2%; Pred. No. 9.2;
Matches 43; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
Qy 10 GGCTCGCAATTGAACCTCACTGAAAGGTTGTCGAGGATTCGTCGATGAGTACCAATG 69
Db 113 GGGCTGCGAGTTGAACCTGGAANATGTTGACGATCTTGGGGACGAACTGTAAAG 172
Qy 70 GCTGTGA 76
Db 173 GTTGCA 179

RESULT 8
BU636714
LOCUS      010F08 Infected Arabidopsis Leaf Arabidopsis thaliana cDNA, mRNA
DEFINITION
ACCESSION BU636714.1 GI:23303969
VERSION    BU636714
KEYWORDS   EST.
SOURCE     Arabidopsis thaliana (thale cress)
ORGANISM   Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE  1 (bases 1 to 656)
AUTHORS    Lundsgaard, M., Emmersen, J., Nielsen, K.L., Wilson, I., Somerville, S.
            and Welinder, K.G.
TITLE      EST sequencing of Erysiphe cichoracearum infected Arabidopsis
JOURNAL    Unpublished (2002)
COMMENT    Contact: Karen G. Welinder
            Institut for bioteknologi
            Aalborg Universitet
            Sohngaardsholmsvej 49, 9000 Aalborg, Denmark
            Tel: +45 96358467
            Fax: +45 98141808
            Email: kgw@bio.auc.dk.
            Location/Qualifiers
            1..656
            /organism="Arabidopsis thaliana"
            /mol_type="mRNA"
            /ecotype="Columbia"
            /db_xref="taxon:3702"
            /dev_stage="Plant 3 weeks old, three days post infection"
            /clone_lib="Infected Arabidopsis Leaf"
            /note="Organ: Leaf; Vector: pBluescript; Mixed cDNA"

```





```

Db      368 AGATGCTTCTCCACTTGAGATCATGATGAAGGCTCTTGAGAGATTGCGAGACCAATCGC 427
QY      66 AATGCTGTGA 76
      ||| ||| |||
Db      428 AATGCTTTTA 438

RESULT 12
BX845438/c
LOCUS   BX845438
DEFINITION   BX845438 NICHD XGC 001 Xenopus laevis cDNA clone IMAGE998P2114225 ;
IMAGE:6641733 5', mRNA sequence.
ACCESSION   BX845438
VERSION     BX845438.1 GI:39735223
KEYWORDS
SOURCE      Xenopus laevis (African clawed frog)
ORGANISM    Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus; Xenopus.
REFERENCE   1 (bases 1 to 711)
AUTHORS     Heil, O., Neubert, P., Peters, M., Radelof, U., Schneider, D.,
Schroth, A., Korn, B. and Landgrebe, J.
TITLE       Xenopus laevis UniGene Set 1 (RZPDLib No. 988)
JOURNAL     Unpublished (2003)
COMMENT     Contact: Ina Rolfs
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany
RZPD; IMAGE998P2114225.
RZPDLib; I.M.A.G.E. cDNA Clone Collection (amp- resistant) (RZPDLib
No. 988) http://www.rzpd.de/cgi-
bin/products/showLib.pl.cgi/response?libNo=988 RZPDLib; Xenopus
laevis UniGene Set 1 (RZPDLib No. 988) http://www.rzpd.de/cgi-
bin/products/showLib.pl.cgi/response?libNo=988 Contact: Ina Rolfs
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Heubnerweg 6, D-14059 Berlin, Germany
Tel: +49 30 32639 101
Fax: +49 30 32639 111
www.rzpd.de
This clone is available royalty-free from RZPD;
contact RZPD (clone@rzpd.de) for further information. Seq primer:
SP6, 5' ATTAGTGACACTATAG 3'.
FEATURES   source
            1..711
            /organism="Xenopus laevis"
            /mol_type="mRNA"
            /db_xref="taxon:8355"
            /clone="IMAGE998P2114225 ; IMAGE:6641733"
            /tissue_type="oocytes"
            /lab_host="DH10B (phage-resistant)"
            /clone_lib="NICHD_XGC_001"
            /notes="Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI;
            Cloned unidirectionally. Primer: Oligo dT. Average insert
            size 2.2 kb. Constructed by Life Technologies."
ORIGIN
Query Match      39.7%; Score 30.2; DB 5; Length 711;
Best Local Similarity 69.5%; Pred. No. 18;
Matches 41; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY      16 CCAATTGAACCTACTGAAAGGTTGTTGAGGAGTTCTGTGATGAAGTACCAATGGCTGT 74
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      210 CCAACAGAACTCACTGAAGAGGACTGGGGGATGTGGGATGGAGCACCAGTGGTTTT 152

RESULT 13
BU915084/c
LOCUS   BU915084
DEFINITION   BU915084 NICHD_XGC_001 Xenopus laevis cDNA clone
IMAGE:6641710 5', mRNA sequence.
ACCESSION   BU915084
VERSION     BU915084.1 GI:24096998
KEYWORDS
SOURCE      Xenopus laevis (African clawed frog)
ORGANISM    Xenopus laevis
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus; Xenopus.
REFERENCE   1 (bases 1 to 800)
AUTHORS     Elgar, G., Clark, M.S., Meek, S., Smith, S., Warner, S., Edwards, Y.J.,
Bouchireb, N., Cottage, A., Yeo, G.S., Umrانيا, Y., Williams, G. and
Brenner, S.
TITLE       Generation and analysis of 25 Mb of genomic DNA from the pufferfish
Fugu rubripes by sequence scanning
JOURNAL     Genome Res. 9 (10), 960-971 (1999)
MEDLINE     99455097
PUBMED      10523524
REFERENCE   2 (bases 1 to 446)
AUTHORS     Elgar, G., Clark, M., Smith, S., Meek, S., Warner, S., Umrانيا, Y.,
Williams, G. and Brenner, S.
TITLE       Direct Submission
JOURNAL     Submitted (08-DEC-1997) MRC Human Genome Mapping Project Resource
Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@hgmrc.ac.uk
Vector: pBluescript II KS

```

```

SOURCE      Xenopus laevis (African clawed frog)
ORGANISM    Xenopus laevis
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus; Xenopus.
REFERENCE   1 (bases 1 to 800)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Martha Rabbert, Steven L. Klein, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LHAM14225 row: O column: 22
High quality sequence stop: 590.
FEATURES   source
            1..800
            /organism="Xenopus laevis"
            /mol_type="mRNA"
            /db_xref="taxon:8355"
            /clone="IMAGE:6641710"
            /tissue_type="oocytes"
            /lab_host="DH10B (phage-resistant)"
            /clone_lib="NICHD_XGC_001"
            /notes="Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI;
            Cloned unidirectionally. Primer: Oligo dT. Average insert
            size 2.2 kb. Constructed by Life Technologies."
ORIGIN
Query Match      39.7%; Score 30.2; DB 5; Length 800;
Best Local Similarity 69.5%; Pred. No. 18;
Matches 41; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY      16 CCAATTGAACCTACTGAAAGGTTGTTGAGGAGTTCTGTGATGAAGTACCAATGGCTGT 74
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      210 CCAACAGAACTCACTGAAGAGGACTGGGGGATGTGGGATGGAGCACCAGTGGTTTT 152

RESULT 14
FR0020459/c
LOCUS   FR0020459
DEFINITION   F. rubripes GSS sequence, clone 041PlidF7, genomic survey sequence.
ACCESSION   AL013342
VERSION     AL013342.1 GI:2679710
KEYWORDS    GSS; genome survey sequence.
SOURCE      Takifugu rubripes (Pugu rubripes)
ORGANISM    Takifugu rubripes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodontidae; Takifugu.
REFERENCE   1
AUTHORS     Elgar, G., Clark, M.S., Meek, S., Smith, S., Warner, S., Edwards, Y.J.,
Bouchireb, N., Cottage, A., Yeo, G.S., Umrانيا, Y., Williams, G. and
Brenner, S.
TITLE       Generation and analysis of 25 Mb of genomic DNA from the pufferfish
Fugu rubripes by sequence scanning
JOURNAL     Genome Res. 9 (10), 960-971 (1999)
MEDLINE     99455097
PUBMED      10523524
REFERENCE   2 (bases 1 to 446)
AUTHORS     Elgar, G., Clark, M., Smith, S., Meek, S., Warner, S., Umrانيا, Y.,
Williams, G. and Brenner, S.
TITLE       Direct Submission
JOURNAL     Submitted (08-DEC-1997) MRC Human Genome Mapping Project Resource
Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@hgmrc.ac.uk
Vector: pBluescript II KS

```

```

ORIGIN
Query Match          39.2%; Score 29.8; DB 5; Length 561;
Best Local Similarity 66.2%; Pred. No. 23;
Matches 43; Conservative 0; Mismatches 22; Indels 0; Gaps 0

Qy      5  CAGACGGCTCGCCAAATTGAACCTCACTGARAAGGTTGTTGAGGAGTTTCGTGGATCAAGTAC 64
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      341 CTGAAGAATCGGCCCAATTTCTTGACTGTTATAGGTTGTAGTAGATTTAGTGGATGATGTAC 282

Qy      65  CAATG 69
      ||| |||
Db      281  TAACG 277

Search completed: January 17, 2005, 16:28:48
Job time : 154.251 secs

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 16, 2005, 21:55:52 ; Search time 17605.5 Seconds  
(without alignments)  
17070.053 Million cell updates/sec

Title: US-09-551-494-5  
Perfect score: 6355  
Sequence: 1 gatgtttaatagtttcga.....taaccgcggtagggccca 6355

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_hcg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6355	100.0	6355	6	AX040177 Sequence
2	6355	100.0	6355	6	AX795380 Sequence
3	6355	100.0	6355	14	M34077 Tobacco mil
4	5951.6	93.7	6356	14	AB078435 Tobacco m
5	2631	41.4	6384	6	E07987 Tobacco mos
6	2631	41.4	6384	14	X02144 Tobacco mos
7	2631	41.4	6385	14	AB083196 Tobacco mo
8	2626.2	41.3	6384	14	AJ132845 tobacco mo
9	2623.8	41.3	6383	14	AJ243571 tobacco m
10	2621.6	41.3	6384	14	AF155507 Tobacco m
11	2620.6	41.2	6383	14	AF332868 Tobacco mo
12	2617.4	41.2	6383	14	292909 Tobacco mos
13	2612.6	41.1	6383	14	AJ417701 Tobacco mo
14	2529.6	39.8	6357	14	AB069853 Pepper mi
15	2528.4	39.8	6357	14	M81413 Nucleotide
16	2528	39.8	6357	14	AB000709 Pepper mi
17	2496.4	39.3	6357	14	AJ308228 Pepper Mi
18	2489.2	39.2	6395	14	D63809 Tobacco mos
19	2479.2	39.0	6395	14	AF395129 Tobacco m

20	2472.8	38.9	6395	14	AF395128	Tobacco m
21	2469.6	38.9	6395	14	AF273221	Tobacco m
22	2469.6	38.9	6395	14	AF395127	Tobacco m
23	2468	38.8	6395	6	AR173320	Sequence
24	2468	38.8	6395	6	AX098414	Sequence
25	2466.4	38.8	6395	6	AR271575	Sequence
26	2466.4	38.8	6395	6	AX040174	Sequence
27	2466.4	38.8	6395	14	TOTMV4	V01408 Tobacco mos
28	2461.6	38.7	6398	14	TOTMV5	V01409 Tobacco mos
29	2460	38.7	6395	14	TMVCG	X68110 Tobacco mos
30	2455.4	38.6	6439	6	AR173321	Sequence
31	2455.4	38.6	6439	6	AX098415	Sequence
32	2453.6	38.6	6395	14	TMV11933	AJ011933 Tobacco m
33	2452.2	38.6	6475	6	AR173323	Sequence
34	2452.2	38.6	6475	6	AX098417	Sequence
35	2452	38.6	6395	6	AX194392	Sequence
36	2452	38.6	6395	14	AF165190	AF165190 Tobacco m
37	2449.8	38.5	6446	6	AR173324	Sequence
38	2449.8	38.5	6446	6	AX098418	Sequence
39	2431.2	38.3	6425	6	AR173322	Sequence
40	2431.2	38.3	6425	6	AX098416	Sequence
41	2370.2	37.3	6524	14	AB089381	AB089381 Paprika m
42	2327.4	36.6	6506	14	MTVNGHYPER	L11665 Obuda peppe
43	2325.8	36.6	6507	14	MTVGRNA	D13438 Obuda peppe
44	2291.4	36.1	11641	6	AX466963	AX466963 Sequence
45	2286.6	36.0	7685	6	BD263877	BD263877 Viral exp

ALIGNMENTS

RESULT 1  
AX040177  
LOCUS AX040177 6355 bp DNA linear PAT 18-NOV-2000  
DEFINITION Sequence 5 from Patent WO0063397.  
ACCESSION AX040177  
VERSION AX040177.1 GI:11230127  
KEYWORDS synthetic construct  
SOURCE synthetic construct  
ORGANISM artificial sequences.  
REFERENCE 1  
AUTHORS Meulewaeter, P., Cornelisse, Jacobs J., van Eldik, G. and Metzlaiff, M.  
TITLE Methods and means for delivering inhibitory rna to plants and appl  
ications thereof  
JOURNAL Patent: WO 0063397-A 5 26-OCT-2000;  
Aventis CropScience N.V. (BE)  
FEATURES  
source  
1..6355  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="cDNA copy of the nucleotide sequence of the genome  
of TMV-U2"

ORIGIN	Query Match	Best Local Similarity	Score	DB 6;	Length
	Matches 6355;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	GATGTTTTTATAGCTTTTCGACACACATATAAACAACACATATTTACAAACACACA	60		
Db	1	GATGTTTTTATAGCTTTTCGACACACATATAAACAACACATATTTACAAACACACA	60		
Qy	61	AACAACAACAAATGGCACACATACATCTATAATTAGCAACGCCCTTCTTGAAGCGTGAG	120		
Db	61	AACAACAACAAATGGCACACATACATCTATAATTAGCAACGCCCTTCTTGAAGCGTGAG	120		
Qy	121	TGGTAAACAACTCTCGTTAATGACCTTGCAGAGCGCATGTACGATACGGCGGTGA	180		
Db	121	TGGTAAACAACTCTCGTTAATGACCTTGCAGAGCGCATGTACGATACGGCGGTGA	180		
Qy	181	AGATTTTAAAGCCCGGACCGTAGACCAAGGTCACTTTTCCAAAACACTATTAGCGAGA	240		

Db 181 AGAATTTAAAGCCCGCGAGCGTAGACCAAGAGTCAACTTTTCCAAACTATTAGCGAAG 240  
QY 241 GCAACGCTTCTAGTCTCTCAACCGGTACCGGAGTTCAGATTACCTTTTATAAATCTCA 300  
Db 241 GCAACGCTTCTAGTCTCTCAACCGGTACCGGAGTTCAGATTACCTTTTATAAATCTCA 300  
QY 301 AATGCGGTACACAGTTTGGCTGGAGTTTGAGAGCATTTAGAAATCTGATGCT 360  
Db 301 AATGCGGTACACAGTTTGGCTGGAGTTTGAGAGCATTTAGAAATCTGATGCT 360  
QY 361 ACAAGTTCCTTAGATCCGACATATGATATAGTGGAACTTTGACGACATTTGTT 420  
Db 361 ACAAGTTCCTTAGATCCGACATATGATATAGTGGAACTTTGACGACATTTGTT 420  
QY 421 CAAAGCAGGATTAAGTGCATTTGCTGATGCCCCAATCTGGACATACAGAGATATAATGAG 480  
Db 421 CAAAGCAGGATTAAGTGCATTTGCTGATGCCCCAATCTGGACATACAGAGATATAATGAG 480  
QY 481 GCACGAGGACAAAGGACTCAATTTGAGATGTTATTTGTCAGANTGCTCGTTCTAAACA 540  
Db 481 GCACGAGGACAAAGGACTCAATTTGAGATGTTATTTGTCAGANTGCTCGTTCTAAACA 540  
QY 541 GGTAAATTCCTGAGTTTCAAGGGAGGCTTTTAAACAGGTATGACAGAGCTCCCAACGAGT 600  
Db 541 GGTAAATTCCTGAGTTTCAAGGGAGGCTTTTAAACAGGTATGACAGAGCTCCCAACGAGT 600  
QY 601 CTGCTGCTTAAACCTTTTCAAGATTCAGATACATCCGACAGAGATAGTGTAGAG 660  
Db 601 CTGCTGCTTAAACCTTTTCAAGATTCAGATACATCCGACAGAGATAGTGTAGAG 660  
QY 661 ATACGCTGTGCTGACAGTTTGTATGATGATTTTCCATTTTGGCAGAGCATTTACT 720  
Db 661 ATACGCTGTGCTGACAGTTTGTATGATGATTTTCCATTTTGGCAGAGCATTTACT 720  
QY 721 AATATCTAAGAAATATACATGATGATGATGATGATGATGATGATGATGATGATGAT 780  
Db 721 AATATCTAAGAAATATACATGATGATGATGATGATGATGATGATGATGATGATGAT 780  
QY 781 AGACGACGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840  
Db 781 AGACGACGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840  
QY 841 TGTCTCTTTTCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900  
Db 841 TGTCTCTTTTCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900  
QY 901 CTTGCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960  
Db 901 CTTGCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960  
QY 961 ATTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020  
Db 961 ATTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020  
QY 1021 TCTGTACAAGAGTGTAGACAAGTAGGCTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAG 1080  
Db 1021 TCTGTACAAGAGTGTAGACAAGTAGGCTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAG 1080  
QY 1081 AGACGCTTTGCTTACAAGAAACCTTGGCCATGTTTCAACATGAAAGAGCAATCTTTAG 1140  
Db 1081 AGACGCTTTGCTTACAAGAAACCTTGGCCATGTTTCAACATGAAAGAGCAATCTTTAG 1140  
QY 1141 AGACAGGCTTCGGTTAACTTTTGGTTCCCTAGATGAGGACATGGTGTAGTACCGCT 1200  
Db 1141 AGACAGGCTTCGGTTAACTTTTGGTTCCCTAGATGAGGACATGGTGTAGTACCGCT 1200  
QY 1201 GTTGTAGGGTCTTATTAACGACAAAAGATGACAAGGAGTGGCTCATTTGTTAATCGTGA 1260  
Db 1201 GTTGTAGGGTCTTATTAACGACAAAAGATGACAAGGAGTGGCTCATTTGTTAATCGTGA 1260  
QY 1261 CTTGCTTTTACAGTGTCTTAATCATATCAGAACATATCAGGCAAAAGCGTTTAACTTACCA 1320

Db 1261 CTTGCTTTTACAGTGTCTTAATCATATCAGAACATATCAGGCAAAAGCGTTTAACTTACCA 1320  
QY 1321 GAAAGTATATCTTTTCTGAGAGTCTATAAGATCCCGGTGATTAATCAATGGTGTACTGCG 1380  
Db 1321 GAAAGTATATCTTTTCTGAGAGTCTATAAGATCCCGGTGATTAATCAATGGTGTACTGCG 1380  
QY 1381 TAGGTCTGAATGGGATGTAGATAAGCAATTTCTTCAACCCCTTGTCAATGACTTTCTTCTT 1440  
Db 1381 TAGGTCTGAATGGGATGTAGATAAGCAATTTCTTCAACCCCTTGTCAATGACTTTCTTCTT 1440  
QY 1441 GCAGACTAAGCTGCTGCTTCAAGACGATATAGTAAATGGGAAAGTTTCGGTCTTTGGA 1500  
Db 1441 GCAGACTAAGCTGCTGCTTCAAGACGATATAGTAAATGGGAAAGTTTCGGTCTTTGGA 1500  
QY 1501 TAAGACCACTTCTGAACTTATTTGGGATGAGGTGGGCAAAATTTTTCGGAAACGTTTTCCT 1560  
Db 1501 TAAGACCACTTCTGAACTTATTTGGGATGAGGTGGGCAAAATTTTTCGGAAACGTTTTCCT 1560  
QY 1561 CACTATCAAAAGAGAGATTGGTGAAGCAGGAAATTTCTGATGTAAAGTGAAGATGCTCTGAA 1620  
Db 1561 CACTATCAAAAGAGAGATTGGTGAAGCAGGAAATTTCTGATGTAAAGTGAAGATGCTCTGAA 1620  
QY 1621 GATCAAGATCCAGATCTGTATGTCTCATGGAAGACAGGTTCTGTAGTGAATACACAA 1680  
Db 1621 GATCAAGATCCAGATCTGTATGTCTCATGGAAGACAGGTTCTGTAGTGAATACACAA 1680  
QY 1681 GTCTGAGAGTTACCGCATCTAGATATCAAGAGGACTTAGAAGAGCTTGAGCAAAATGTA 1740  
Db 1681 GTCTGAGAGTTACCGCATCTAGATATCAAGAGGACTTAGAAGAGCTTGAGCAAAATGTA 1740  
QY 1741 CGACGCGTTATCAGAAATTTCTATCTTAAAGGCTGTGATAATTTTCGATATCCGAAAGTT 1800  
Db 1741 CGACGCGTTATCAGAAATTTCTATCTTAAAGGCTGTGATAATTTTCGATATCCGAAAGTT 1800  
QY 1801 CAAAGCATGTGCAAGGCTTTAGATGTTAGTCTGTAGTGGCAGCAGAGTAAATCGTTGC 1860  
Db 1801 CAAAGCATGTGCAAGGCTTTAGATGTTAGTCTGTAGTGGCAGCAGAGTAAATCGTTGC 1860  
QY 1861 AGTGGCCGAGAAATAGAAAGCGGTTTAACTCTTCTTTTGTATGAAAGCGAGAGTGT 1920  
Db 1861 AGTGGCCGAGAAATAGAAAGCGGTTTAACTCTTCTTTTGTATGAAAGCGAGAGTGT 1920  
QY 1921 GGTAAAGGCTTTTAAAGCAACGCGCTCTGAGCGCGTGTATGTTTGAACCGACATCCGA 1980  
Db 1921 GGTAAAGGCTTTTAAAGCAACGCGCTCTGAGCGCGTGTATGTTTGAACCGACATCCGA 1980  
QY 1981 AGAGGTGAACGTAAATAAATTTTCTATGCTGAGAAAGGAGATTGCGTGTGTGTCAGA 2040  
Db 1981 AGAGGTGAACGTAAATAAATTTTCTATGCTGAGAAAGGAGATTGCGTGTGTGTCAGA 2040  
QY 2041 AAGTCAATGTTTGAAGATGCTAACTTAGAGCAACGAGGATTTGGAGTCCCTCAACGATTT 2100  
Db 2041 AAGTCAATGTTTGAAGATGCTAACTTAGAGCAACGAGGATTTGGAGTCCCTCAACGATTT 2100  
QY 2101 CCATTAAGGCTTCGGTGGATGTGATTAACAAGCAATGGCATCGGTTGTCTACATGG 2160  
Db 2101 CCATTAAGGCTTCGGTGGATGTGATTAACAAGCAATGGCATCGGTTGTCTACATGG 2160  
QY 2161 CTCACCTCAAGTTCAACAAATGAAGAACTATGTCGACAGTTTGGCAGCTTGTGTCGCGC 2220  
Db 2161 CTCACCTCAAGTTCAACAAATGAAGAACTATGTCGACAGTTTGGCAGCTTGTGTCGCGC 2220  
QY 2221 CACTGTATCAAAATCTATGCAAGTCACTAAAGGATGAACTCGGGTATGATTTCTGATCCAG 2280  
Db 2221 CACTGTATCAAAATCTATGCAAGTCACTAAAGGATGAACTCGGGTATGATTTCTGATCCAG 2280  
QY 2281 GGAGAAAGTGGTGGTGGGATGCTTCAAAAGTGGCTCTCTCAAACTCGGCGCAA 2340  
Db 2281 GGAGAAAGTGGTGGTGGGATGCTTCAAAAGTGGCTCTCTCAAACTCGGCGCAA 2340  
QY 2341 AGGTCAATTCATGGGAGTTGCTCTGGATTAACAAGGGGAAATGTTTACTGCACCTTCTATC 2400  
Db 2341 AGGTCAATTCATGGGAGTTGCTCTGGATTAACAAGGGGAAATGTTTACTGCACCTTCTATC 2400







Matches 6355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	GATGTTTTTAATAGTTTTGCAACAACAACAATTAATAAACAACAACAACAATTAATAACAACA	60
Db	1	GATGTTTTTAATAGTTTTGCAACAACAACAATTAATAAACAACAACAACAATTAATAACAACA	60
Qy	61	AACAAACAACAAATGGCACACATACAATCTATAAATAGCAAGCGCCCTCTTGAAGACGGTGAG	120
Db	61	AACAAACAACAAATGGCACACATACAATCTATAAATAGCAAGCGCCCTCTTGAAGACGGTGAG	120
Qy	121	TGGTAAAAACACTCTCTGTTTAATGACCTTGTGCAAGAGCGGATGATACGATACGGCCGTGGA	180
Db	121	TGGTAAAAACACTCTCTGTTTAATGACCTTGTGCAAGAGCGGATGATACGATACGGCCGTGGA	180
Qy	181	AGAAATTAACCGCCGACCGGTAGACCAAGAGTCAACTTTCCAAACACTATTAGCGNAGA	240
Db	181	AGAAATTAACCGCCGACCGGTAGACCAAGAGTCAACTTTCCAAACACTATTAGCGNAGA	240
Qy	241	GCAAAACGCTTCTAGTCTCCAAACGGTACCGGAGTTCAGATTACCTTTTATAATACTCA	300
Db	241	GCAAAACGCTTCTAGTCTCCAAACGGTACCGGAGTTCAGATTACCTTTTATAATACTCA	300
Qy	301	AAATGCGGTAACAAGTTTGGCTGGAGTTTGGAGGCAATTAGAAATTTGGAATPATCTGATGCT	360
Db	301	AAATGCGGTAACAAGTTTGGCTGGAGTTTGGAGGCAATTAGAAATTTGGAATPATCTGATGCT	360
Qy	361	ACAAAGTTCCCTATGGATCGCGGACATATGATATAGTGGGAACTTTTCGACACATTTGTT	420
Db	361	ACAAAGTTCCCTATGGATCGCGGACATATGATATAGTGGGAACTTTTCGACACATTTGTT	420
Qy	421	CAAAAGGCAAGGATTAACGTGCAATGCTGATGCCCAATCTGCAATACAGATATATATGAG	480
Db	421	CAAAAGGCAAGGATTAACGTGCAATGCTGATGCCCAATCTGCAATACAGATATATATGAG	480
Qy	481	GCAAGAGGCAAAAGGACTCAATTTGAGATGTATTTGTCCAGATTTGCTCGTTCTAAACA	540
Db	481	GCAAGAGGCAAAAGGACTCAATTTGAGATGTATTTGTCCAGATTTGCTCGTTCTAAACA	540
Qy	541	GGTAATTCCTGAGTTTCAAGGGAGGCTTTTAACAGSTATGCAAGCTCCCAACGAGT	600
Db	541	GGTAATTCCTGAGTTTCAAGGGAGGCTTTTAACAGSTATGCAAGCTCCCAACGAGT	600
Qy	601	CTGCTGCTCTAAACTTTTCAAGGATTCGCAATACATCCGCCAGAGAAATGCTGATAGAG	660
Db	601	CTGCTGCTCTAAACTTTTCAAGGATTCGCAATACATCCGCCAGAGAAATGCTGATAGAG	660
Qy	661	ATACGCTGTTGCTGTCACAGTTTGTATGATATTCCTGTGTCATGAGTTTGGAGCTGCGTT	720
Db	661	ATACGCTGTTGCTGTCACAGTTTGTATGATATTCCTGTGTCATGAGTTTGGAGCTGCGTT	720
Qy	721	AATATCTAAGAAATATACATGATATGATATGATATTCCTGTGTCATGAGTTTGGAGCTGCGTT	780
Db	721	AATATCTAAGAAATATACATGATATGATATGATATTCCTGTGTCATGAGTTTGGAGCTGCGTT	780
Qy	781	AGACCCAGCGGTTACGCTTAATGAATAGGCGCAACTTTTCAAAAGAGAGGATGATGA	840
Db	781	AGACCCAGCGGTTACGCTTAATGAATAGGCGCAACTTTTCAAAAGAGAGGATGATGA	840
Qy	841	TGTTTCTTTTCTTCTGCTGATGAAGTACTTTTAAATTTAGTGCATATAAATACAAAATAT	900
Db	841	TGTTTCTTTTCTTCTGCTGATGAAGTACTTTTAAATTTAGTGCATATAAATACAAAATAT	900
Qy	901	CTTGCAATATGATAGTAAATCTTACTTTCTGCTTCTAGTAGAATAGTTTACTTTAAGGA	960
Db	901	CTTGCAATATGATAGTAAATCTTACTTTCTGCTTCTAGTAGAATAGTTTACTTTAAGGA	960
Qy	961	ATTTTGTAGTCACTAGGTTTAACTTGGTTTGTAAATTTTACCAAGTAGATACCTATAT	1020
Db	961	ATTTTGTAGTCACTAGGTTTAACTTGGTTTGTAAATTTTACCAAGTAGATACCTATAT	1020
Qy	1021	TCTGTACAGAGTGTTAGACAAGTGGGTGTAGTGTATGATCAGTTCTATGAGGCGATGGA	1080
Db	1021	TCTGTACAGAGTGTTAGACAAGTGGGTGTAGTGTATGATCAGTTCTATGAGGCGATGGA	1080

Qy	1081	AGACGCTTTGCTTACAGAAAAACCTTGGCCATGTTTCAACACATGAAAGAGCAATCTTTAG	1140
Db	1081	AGACGCTTTGCTTACAGAAAAACCTTGGCCATGTTTCAACACATGAAAGAGCAATCTTTAG	1140
Qy	1141	AGACACGGCTTCGGTTAACTTTTGGTTCCCTAAAGATGAAGGACATGGTGTATGATACCGCT	1200
Db	1141	AGACACGGCTTCGGTTAACTTTTGGTTCCCTAAAGATGAAGGACATGGTGTATGATACCGCT	1200
Qy	1201	GTTTGGGGTTCTATTATCCAGGAAAAAAGATGACAGAGGTGAGGTCAATGTTTAAATCGTGA	1260
Db	1201	GTTTGGGGTTCTATTATCCAGGAAAAAAGATGACAGAGGTGAGGTCAATGTTTAAATCGTGA	1260
Qy	1261	CTTTCGTTTACACAGTGTCTTAATCATATCAGAACATATCAAGCCAAAGCGTTTAACTTACCA	1320
Db	1261	CTTTCGTTTACACAGTGTCTTAATCATATCAGAACATATCAAGCCAAAGCGTTTAACTTACCA	1320
Qy	1321	GAAACGTTATTCTTTCTGCGAGTCTATTAAGATCCCGGTGATTAATCAATGGTGTACTGTC	1380
Db	1321	GAAACGTTATTCTTTCTGCGAGTCTATTAAGATCCCGGTGATTAATCAATGGTGTACTGTC	1380
Qy	1381	TAGGTCGTGAATGGGATGTAGATAAGCAATTTCTTCAACCTTGTCAATGACTTTCTTCTTT	1440
Db	1381	TAGGTCGTGAATGGGATGTAGATAAGCAATTTCTTCAACCTTGTCAATGACTTTCTTCTTT	1440
Qy	1441	GCAGACTTAAGCTGCGCTTCAAGACGATATAGTAATGGGAAAGTTTCGGTCTTGGGA	1500
Db	1441	GCAGACTTAAGCTGCGCTTCAAGACGATATAGTAATGGGAAAGTTTCGGTCTTGGGA	1500
Qy	1501	TAAGACCACTTCTGAACTTATTTGGGATGAGGTGGGCAAAATTTTGGGAAACGTTTTCCC	1560
Db	1501	TAAGACCACTTCTGAACTTATTTGGGATGAGGTGGGCAAAATTTTGGGAAACGTTTTCCC	1560
Qy	1561	CACATCAAGAGAGAGATTTGGTGAGCAGGAAAAATTTCTGGATGTAAAGTGAGAAATGCTCGAA	1620
Db	1561	CACATCAAGAGAGAGATTTGGTGAGCAGGAAAAATTTCTGGATGTAAAGTGAGAAATGCTCGAA	1620
Qy	1621	GATCAAGATCCCAGATCTGATGTCATCATGGAAGACAGGTTCTGTAGCTGAATACACCAA	1680
Db	1621	GATCAAGATCCCAGATCTGATGTCATCATGGAAGACAGGTTCTGTAGCTGAATACACCAA	1680
Qy	1681	GTCTGAGGAGTTAACCCGATCTAGATATCAAGAAAGGACTTAGAAGAAAGCTGAGCAAAATGTA	1740
Db	1681	GTCTGAGGAGTTAACCCGATCTAGATATCAAGAAAGGACTTAGAAGAAAGCTGAGCAAAATGTA	1740
Qy	1741	CGACGGTTATCAGAAATPATCTATCTTAAAGGTGCTGATAATTCGATATCGGGAAGTT	1800
Db	1741	CGACGGTTATCAGAAATPATCTATCTTAAAGGTGCTGATAATTCGATATCGGGAAGTT	1800
Qy	1801	CAAGACATGTGCAAGGCTTTAGATGTTAGTCTCTGATGTGGCAGCAGAGTAATCGTTGC	1860
Db	1801	CAAGACATGTGCAAGGCTTTAGATGTTAGTCTCTGATGTGGCAGCAGAGTAATCGTTGC	1860
Qy	1861	AGTGGCGGAGAAATAGAAGCGGTTTAACTCTTACTTTTGAAGCAACCGAGGAGAAATGT	1920
Db	1861	AGTGGCGGAGAAATAGAAGCGGTTTAACTCTTACTTTTGAAGCAACCGAGGAGAAATGT	1920
Qy	1921	GGCTAAAGGCTCTTAAAGGACACGCGCTCTGAGGCGGTGATGTCTTGAAACCGACATCCGA	1980
Db	1921	GGCTAAAGGCTCTTAAAGGACACGCGCTCTGAGGCGGTGATGTCTTGAAACCGACATCCGA	1980
Qy	1981	AGAGGTGAACGTAATAATAATTTCTATTGCTGAGAAAGGAGGATTTGCTGTGTGTCAGA	2040
Db	1981	AGAGGTGAACGTAATAATAATTTCTATTGCTGAGAAAGGAGGATTTGCTGTGTGTCAGA	2040
Qy	2041	AAGTCATGTTTTCACGAATGCTAACTTAGAGCACCGAGGAGTTGGAGTCCCTCAACGATTT	2100
Db	2041	AAGTCATGTTTTCACGAATGCTAACTTAGAGCACCGAGGAGTTGGAGTCCCTCAACGATTT	2100
Qy	2101	CCATAAGGCTTGGGTGGATAGTGTGATTAACAAAGCAAAATGCGCATCGGTTGCTACACTGG	2160
Db	2101	CCATAAGGCTTGGGTGGATAGTGTGATTAACAAAGCAAAATGCGCATCGGTTGCTACACTGG	2160

Qy	2161	CTCACTCAAGTTCACAAATGAAGAACTATGTGACAGATTTGGCAGCTTCGTTGTCGCG	2220
Db	2161	CTCACTCAAGTTCACAAATGAAGAACTATGTGACAGATTTGGCAGCTTCGTTGTCGCG	2220
Qy	2221	CACGTATCAAACTATGCAAGTCACTAAAGGATGAAGTCGGGTATGATCTGATTCAG	2280
Db	2221	CACGTATCAAACTATGCAAGTCACTAAAGGATGAAGTCGGGTATGATCTGATTCAG	2280
Qy	2281	GGAGAAAGTTCGTTGGGATGTCACTTTGAAAGAGTGGCTCTCAAACTCGCGGCAA	2340
Db	2281	GGAGAAAGTTCGTTGGGATGTCACTTTGAAAGAGTGGCTCTCAAACTCGCGGCAA	2340
Qy	2341	AGGTCATTCATGGGGAGTTCCTGGATTACAGGGGAAATGTTACTGACATTCATC	2400
Db	2341	AGGTCATTCATGGGGAGTTCCTGGATTACAGGGGAAATGTTACTGACATTCATC	2400
Qy	2401	TTATGAAGGAGATAGAAATGGTGACTGAGAGCGACTGGAGAGGGTGGCTGATCATCTGA	2460
Db	2401	TTATGAAGGAGATAGAAATGGTGACTGAGAGCGACTGGAGAGGGTGGCTGATCATCTGA	2460
Qy	2461	TACAATGGTATATCTGATATTCGAAAGCTCCAAAATCTGAGGAAACAAATGAGAGCGG	2520
Db	2461	TACAATGGTATATCTGATATTCGAAAGCTCCAAAATCTGAGGAAACAAATGAGAGCGG	2520
Qy	2521	TGAACCCCAACCACTACTGCAAGATGGTACTTGTGGATGGGGTGGCTGGTTGGAAA	2580
Db	2521	TGAACCCCAACCACTACTGCAAGATGGTACTTGTGGATGGGGTGGCTGGTTGGAAA	2580
Qy	2581	GTACAAAGGAGATTTGAAAGATTTGATCTTGATGAGGATTTGATCTTGGTTCTCGAAA	2640
Db	2581	GTACAAAGGAGATTTGAAAGATTTGATCTTGATGAGGATTTGATCTTGGTTCTCGAAA	2640
Qy	2641	ACAAGCTGCTATGATCAGAAGAGGGCTAAATTCATCTGGAATGATAAGAGCCAAAT	2700
Db	2641	ACAAGCTGCTATGATCAGAAGAGGGCTAAATTCATCTGGAATGATAAGAGCCAAAT	2700
Qy	2701	GGACAAATGTGAGAACGGTAGATTCATCTTAATGCATCCAAAACCGGATCACAAAG	2760
Db	2701	GGACAAATGTGAGAACGGTAGATTCATCTTAATGCATCCAAAACCGGATCACAAAG	2760
Qy	2761	GCCTTTTATGATGAAGGTTGATGCTGCACACCGTTGTTGTTAACTTCCTGGTGCCTAT	2820
Db	2761	GCCTTTTATGATGAAGGTTGATGCTGCACACCGTTGTTGTTAACTTCCTGGTGCCTAT	2820
Qy	2821	CTCTGGTTCGACATCGCATACATTTACGGAGATACACAGCAGATTCCTTTCAATTAAC	2880
Db	2821	CTCTGGTTCGACATCGCATACATTTACGGAGATACACAGCAGATTCCTTTCAATTAAC	2880
Qy	2881	AGTTCAGAAATTCCTGGTATCCCAACATTTTGAGAACCTGCAAGTGGATGAAGTTGAGAT	2940
Db	2881	AGTTCAGAAATTCCTGGTATCCCAACATTTTGAGAACCTGCAAGTGGATGAAGTTGAGAT	2940
Qy	2941	GAGGAGACACACTGAGATGCCCAGGTGATGTAATTTTCTCAATCGAAGTACGA	3000
Db	2941	GAGGAGACACACTGAGATGCCCAGGTGATGTAATTTTCTCAATCGAAGTACGA	3000
Qy	3001	AGGACGGGTGACACCACTTCAATCTGTAACAGATCGGTCTCATCTGAGATGATAGCGG	3060
Db	3001	AGGACGGGTGACACCACTTCAATCTGTAACAGATCGGTCTCATCTGAGATGATAGCGG	3060
Qy	3061	TAAGGGAGTACTAAACAGTGTTCCTCAACCACTTAAGAGGGGAAATTTGTAATCTTCACTCA	3120
Db	3061	TAAGGGAGTACTAAACAGTGTTCCTCAACCACTTAAGAGGGGAAATTTGTAATCTTCACTCA	3120
Qy	3121	GGCTGATAAATTTGAGTTAGAGGAGAGGGCTATAAGAAATGTGAACCGGTTCAATGAGAT	3180
Db	3121	GGCTGATAAATTTGAGTTAGAGGAGAGGGCTATAAGAAATGTGAACCGGTTCAATGAGAT	3180
Qy	3181	CCAAGGAGAAACCTTTGAAGATGTGCTGGTCAGATTCAGCGCAACTCCACTGACTCT	3240
Db	3181	CCAAGGAGAAACCTTTGAAGATGTGCTGGTCAGATTCAGCGCAACTCCACTGACTCT	3240
Qy	3241	GATTTCCAGTCTTCCCGCATGTTCTTAGTCGCTCTGACTAGACACAAAAGAGCTTCAA	3300

Db	3341	GATTTCCAGTCTTCCCGCATGTTCTAGTCGCTCTGACTAGACACAAAAGAGCTTCAA	3300
Qy	3301	ATATTACACCGTAGTGTAGATCCTTTAGTACAGATAATTTAGTGATTTGCTTCTTTAAG	3360
Db	3301	ATATTACACCGTAGTGTAGATCCTTTAGTACAGATAATTTAGTGATTTGCTTCTTTAAG	3360
Qy	3361	CTCCTTCTTTTGAATAATATATGTGTAAGCAGGTAGTAGAGCAATTTACAGATGGA	3420
Db	3361	CTCCTTCTTTTGAATAATATATGTGTAAGCAGGTAGTAGAGCAATTTACAGATGGA	3420
Qy	3421	TGCAGTGTTCAAAGGTCAATAATCTCTTTTGTGGCAACACCTAAATCAGAGACTTTCAGA	3480
Db	3421	TGCAGTGTTCAAAGGTCAATAATCTCTTTTGTGGCAACACCTAAATCAGAGACTTTCAGA	3480
Qy	3481	TCTACAGTCTTATTTACAGTATGTCCTCCCTGGTAAATAGTACTATCTTAAACAAGTATGA	3540
Db	3481	TCTACAGTCTTATTTACAGTATGTCCTCCCTGGTAAATAGTACTATCTTAAACAAGTATGA	3540
Qy	3541	TGCTGTTACCATGAGGTTACGTGATATAGTCTTAAATGTGAAGATTTGTCTTGTGATTT	3600
Db	3541	TGCTGTTACCATGAGGTTACGTGATATAGTCTTAAATGTGAAGATTTGTCTTGTGATTT	3600
Qy	3601	TTCCAAAAGTATTTCCGATGCGAAGGAGGTGAACCAATGCTAGAGCCAGTTTTCGCTAC	3660
Db	3601	TTCCAAAAGTATTTCCGATGCGAAGGAGGTGAACCAATGCTAGAGCCAGTTTTCGCTAC	3660
Qy	3661	CGCGGCGGAACCGCAAGGGCTGCAGGACTACTCGAAAAATCTGGTTGCAATGATTAAGAAG	3720
Db	3661	CGCGGCGGAACCGCAAGGGCTGCAGGACTACTCGAAAAATCTGGTTGCAATGATTAAGAAG	3720
Qy	3721	AAATTTCAACCGCACAGACTGACGGGACGATGACATTTGAGAGCAACCGCATCTGTGT	3780
Db	3721	AAATTTCAACCGCACAGACTGACGGGACGATGACATTTGAGAGCAACCGCATCTGTGT	3780
Qy	3781	AGTAGATAAGTATTTGATAGCTATTTTATTAAGAAAGAAATACACAAAATATATTC	3840
Db	3781	AGTAGATAAGTATTTGATAGCTATTTTATTAAGAAAGAAATACACAAAATATATTC	3840
Qy	3841	TGCAGTGTATGCAAGGATTCATGATGATGATGGTTGGAAGAAAGAGTACTATTT	3900
Db	3841	TGCAGTGTATGCAAGGATTCATGATGATGATGGTTGGAAGAAAGAGTACTATTT	3900
Qy	3901	GGACGACTTGGCTAACTCAATTTTACAGATCTGCGGGCATCGATCAGTACAAGCACAT	3960
Db	3901	GGACGACTTGGCTAACTCAATTTTACAGATCTGCGGGCATCGATCAGTACAAGCACAT	3960
Qy	3961	GATCAGGCTCAACCAAAACAGAAATTTGACCTTTCAATTCAGAAATACCTGCTCT	4020
Db	3961	GATCAGGCTCAACCAAAACAGAAATTTGACCTTTCAATTCAGAAATACCTGCTCT	4020
Qy	4021	GCAAAACAATTTCTACCATTCGAAGCAGATCAACGGTATTTTGGCGGTTTCTCAGAGCT	4080
Db	4021	GCAAAACAATTTCTACCATTCGAAGCAGATCAACGGTATTTTGGCGGTTTCTCAGAGCT	4080
Qy	4081	TACAAGTGTCTGCTCAGGCAATTTGATTTCTAAAGAGTTTCTTTTCTTTACTAGGAAAC	4140
Db	4081	TACAAGTGTCTGCTCAGGCAATTTGATTTCTAAAGAGTTTCTTTTCTTTACTAGGAAAC	4140
Qy	4141	TCAGAAACAGATTCAGAAATTTTCTCGGATCTCGACTCGCACTGCTATGATGTGTT	4200
Db	4141	TCAGAAACAGATTCAGAAATTTTCTCGGATCTCGACTCGCACTGCTATGATGTGTT	4200
Qy	4201	AGAACTGGATATTTCTAAGTATGATAAGTACAGAAACGAGTTTCATTTGCTGTAGAGTA	4260
Db	4201	AGAACTGGATATTTCTAAGTATGATAAGTACAGAAACGAGTTTCATTTGCTGTAGAGTA	4260
Qy	4261	TGAATATGGAAGAGTGGGTCTCAATGAGTTTGTGGCGAAGTGTGGAACAAGGGCA	4320
Db	4261	TGAATATGGAAGAGTGGGTCTCAATGAGTTTGTGGCGAAGTGTGGAACAAGGGCA	4320
Qy	4321	CAGGAAAACAATTTGAAGATTTACATTTGCTGGAATTCAGACATGCTGTGGTATCAAG	4380

4321 CAGGAAACCACTTTGAAGGATTACATTTGCTGGAATCAAGACATGTCCTGTGGTATCAAG 4380  
4381 GAAAAGCGGTGATGATGACTTTCATCGGCAATACATGTTATATATAGCAGCTTGTCTGGG 4440  
4381 GAAAAGCGGTGATGATGACTTTCATCGGCAATACATGTTATATATAGCAGCTTGTCTGGG 4440  
4441 TTCAATGTTACCGATGAAAAGGTCATAAAAGGTGCTTTTGTGAGACAGCATTCGCTTTT 4500  
4441 TTCAATGTTACCGATGAAAAGGTCATAAAAGGTGCTTTTGTGAGACAGCATTCGCTTTT 4500  
4501 GTATTTTCCAAAGGTTTGGATTTCCCTGACATTCAGTCATGCTAATCTCATGTGGAA 4560  
4501 GTATTTTCCAAAGGTTTGGATTTCCCTGACATTCAGTCATGCTAATCTCATGTGGAA 4560  
4561 TTTTGGAGCCAACTGTATAGAAAGAGTACGGTTACTTTTGTGGTAGATACATCATACA 4620  
4561 TTTTGGAGCCAACTGTATAGAAAGAGTACGGTTACTTTTGTGGTAGATACATCATACA 4620  
4621 CCATGATAAGGAGCAATAGTATATATGATCTCTTGAAGTTGATCTCCAAACTTGGGGC 4680  
4621 CCATGATAAGGAGCAATAGTATATATGATCTCTTGAAGTTGATCTCCAAACTTGGGGC 4680  
4681 AAAACATATCAAGGATTATGATCACTTAGAAGAGTTAAGGGTGTCTTTGTGCGATGTTGC 4740  
4681 AAAACATATCAAGGATTATGATCACTTAGAAGAGTTAAGGGTGTCTTTGTGCGATGTTGC 4740  
4741 TTGTTCCCTCGAAAACCTGTTAGGCTTTCCGACGTGAAACGACATCAAGGAGGT 4800  
4741 TTGTTCCCTCGAAAACCTGTTAGGCTTTCCGACGTGAAACGACATCAAGGAGGT 4800  
4801 TCATAAACCGGATTCAGTGGTGTGCTTTTAAATGTTGTTAAACAAATTTTGTGTGA 4860  
4801 TCATAAACCGGATTCAGTGGTGTGCTTTTAAATGTTGTTAAACAAATTTTGTGTGA 4860  
4861 TAAATTTTATTTAGAACTTTGTTTTTAAATGCTGTTAGTCTCAGAGATACGTCAAAA 4920  
4861 TAAATTTTATTTAGAACTTTGTTTTTAAATGCTGTTAGTCTCAGAGATACGTCAAAA 4920  
4921 TTAGCGAGTTCATTTGATCTTTGAAAACAGGATGAGATCTCCGGCATTCATGACTAAG 4980  
4921 TTAGCGAGTTCATTTGATCTTTGAAAACAGGATGAGATCTCCGGCATTCATGACTAAG 4980  
4981 TCAGAGTGTAGAAATACACTGTGACCAAGATTATGCTGTTTAAAGATGATGATCTTTT 5040  
4981 TCAGAGTGTAGAAATACACTGTGACCAAGATTATGCTGTTTAAAGATGATGATCTTTT 5040  
5041 CTGATGTAGATTTTACTTTAAAGGTGTTAAGTTAGTTAAGAAAGGATGATGCTGTAGCTG 5100  
5041 CTGATGTAGATTTTACTTTAAAGGTGTTAAGTTAGTTAAGAAAGGATGATGCTGTAGCTG 5100  
5101 ATTTGGTAGTCTGCGGAGTGGAACTCCCGGATAACTCCCGTGGTGGTGTGAGTGTCTT 5160  
5101 ATTTGGTAGTCTGCGGAGTGGAACTCCCGGATAACTCCCGTGGTGGTGTGAGTGTCTT 5160  
5161 GTATTGTAGATAGAGAAATGAAAAGGATGAAAGCAACGCTGGGTGGTATCACGCC 5220  
5161 GTATTGTAGATAGAGAAATGAAAAGGATGAAAGCAACGCTGGGTGGTATCACGCC 5220  
5221 CTGCTTCAAAAAGAAATTTTCTTTTAAAGCTTAATCCCTTAATTTTCAATAACATCCGAGG 5280  
5221 CTGCTTCAAAAAGAAATTTTCTTTTAAAGCTTAATCCCTTAATTTTCAATAACATCCGAGG 5280  
5281 ATGCTGAGAAAGCAACCGTGGCAAGTGTAGTGAATATCAAGAGGATGGCTATGGAAGAG 5340  
5281 ATGCTGAGAAAGCAACCGTGGCAAGTGTAGTGAATATCAAGAGGATGGCTATGGAAGAG 5340  
5341 GATACCTGCTTTATCTTTGGAGTTCGTTTCAATTTGTGTAGTACATAAAAAATAGTAA 5400  
5341 GATACCTGCTTTATCTTTGGAGTTCGTTTCAATTTGTGTAGTACATAAAAAATAGTAA 5400  
5401 GAAAAGGTTTGGGGAACGTTATTTTGTAGTGTGACAGCGCTCGCCCAATTTGAATCACTG 5460  
5401 GAAAAGGTTTGGGGAACGTTATTTTGTAGTGTGACAGCGCTCGCCCAATTTGAATCACTG 5460

5461 AAAAGGTTTGTGAGGATTCGTGATGAAAGTACCAATGGCTGTGAACTCGAAAAGGTTTC 5520  
5461 AAAAGGTTTGTGAGGATTCGTGATGAAAGTACCAATGGCTGTGAACTCGAAAAGGTTTC 5520  
5521 CGGAAAACAAAAGAAAGTAAATAGTAAATTAATAAGAAATTAATAACAGTGT 5580  
5521 CGGAAAACAAAAGAAAGTAAATAGTAAATTAATAAGAAATTAATAACAGTGT 5580  
5581 GTAAAGAGGTTTAAAAATTTAGGAAATTTGAGGATAATGTAAGTGAAGAGGCTATCG 5640  
5641 CGTCATCGAGTACGTTTAAATCAATATGCTTAAATCAATCAATCTCTCCGAGCCAAATTTGT 5700  
5641 CGTCATCGAGTACGTTTAAATCAATATGCTTAAATCAATCAATCTCTCCGAGCCAAATTTGT 5700  
5701 TTACTTATCTTCCGCTTTACGAGATCTGTGAGCTGATCAATCTGTGTACAAATGCAAT 5760  
5701 TTACTTATCTTCCGCTTTACGAGATCTGTGAGCTGATCAATCTGTGTACAAATGCAAT 5760  
5761 GGATAACCAAGTTTCAACGCAACAAAGCTAGGACAAAGCTCAACAGCAATTTGCGGATGC 5820  
5761 GGATAACCAAGTTTCAACGCAACAAAGCTAGGACAAAGCTCAACAGCAATTTGCGGATGC 5820  
5821 CTGAAAACCTGTGCTTGTATGACAGTGAATTTCTTGCATCGGATTTCTATGTGTATAG 5880  
5821 CTGAAAACCTGTGCTTGTATGACAGTGAATTTCTTGCATCGGATTTCTATGTGTATAG 5880  
5881 ATATAATTTGAGCTTGTGATCGGCTTATTAATAAGCTTATTAATAAGCTTATTAATAAG 5940  
5881 ATATAATTTGAGCTTGTGATCGGCTTATTAATAAGCTTATTAATAAGCTTATTAATAAG 5940  
5941 TAGAATAATAGAGTTGATAATCAACCCGACCGAATACCTGAAATCGTTAAACGCGAC 6000  
5941 TAGAATAATAGAGTTGATAATCAACCCGACCGAATACCTGAAATCGTTAAACGCGAC 6000  
6001 TCAGAGGTAGACGATGCTACTAGCTATAGGGCTTCAATCAATTAATTTGGCTAATGA 6060  
6001 TCAGAGGTAGACGATGCTACTAGCTATAGGGCTTCAATCAATTAATTTGGCTAATGA 6060  
6061 ACTGGTTTGTGGAACCTGGCATGTTCAATCAAGCAGGCTTTGAGACTGCTAGTGGACTTGT 6120  
6061 ACTGGTTTGTGGAACCTGGCATGTTCAATCAAGCAGGCTTTGAGACTGCTAGTGGACTTGT 6120  
6121 CTGGAACCAACTCCGCTACTTAGCTATTTGTTGTGAGATTTCTTAAATTAAGTCCGCTG 6180  
6121 CTGGAACCAACTCCGCTACTTAGCTATTTGTTGTGAGATTTCTTAAATTAAGTCCGCTG 6180  
6181 AGACTTTAAATTCAGGGTGGCTGATACCAAAATCAGCAGTGGTGTTCCTCACTTAAA 6240  
6181 AGACTTTAAATTCAGGGTGGCTGATACCAAAATCAGCAGTGGTGTTCCTCACTTAAA 6240  
6241 TATAAGATTTGTCATCTGATCCCAAGTAAACCATGTGATGTTGATGTTGATGTTGATG 6300  
6241 TATAAGATTTGTCATCTGATCCCAAGTAAACCATGTGATGTTGATGTTGATGTTGATG 6300  
6301 TGCGGTAAACATCGGAGAGTTTCGAATCTCCCTTAAACCGCGGTAGCGGCCCA 6355  
6301 TGCGGTAAACATCGGAGAGTTTCGAATCTCCCTTAAACCGCGGTAGCGGCCCA 6355

## RESULT 3

LOCUS TMGCG 6355 bp ss-RNA linear VRL 03-AUG-1993  
DEFINITION Tobacco mild green mosaic virus complete genome.  
ACCESSION M34077.1  
VERSION M34077.1 GI:335243  
KEYWORDS coat protein.  
SOURCE Tobacco mild green mosaic virus  
ORGANISM Tobacco mild green mosaic virus  
REFERENCE 1 (bases 6127 to 6355) Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.









```
Qy 4861 TAAATTTTATTTAGAACTTTGTTTTTAAATGGCTGTAGTCTCAGAGATACCTGTCAAA 4920
Db 4861 TAAATTTTATTTAGAACTTTGTTTTTAAATGGCTGTAGTCTCAGAGATACCTGTCAAA 4920
Qy 4921 TTACGAGTTCATTTGATCTTTTCGAAACAGGATGAGATACCTCCGGCATTCATGACTAAGG 4980
Db 4921 TTACGAGTTCATTTGATCTTTTCGAAACAGGATGAGATACCTCCGGCATTCATGACTAAGG 4980
Qy 4981 TCAAGAGTGTAGAAATATCGACTGTGGACAGATATATGGCTGTGTTAAGAAATGATAGTCTTT 5040
Db 4981 TCAAGAGTGTAGAAATATCGACTGTGGACAGATATATGGCTGTGTTAAGAAATGATAGTCTTT 5040
Qy 5041 CTGATGTAGATTTACTTTAAAGGTGTTAAAGTGTAGTTAAGAAAGGGTATGTGTGCTTAGCTG 5100
Db 5041 CTGATGTAGATTTACTTTAAAGGTGTTAAAGTGTAGTTAAGAAAGGGTATGTGTGCTTAGCTG 5100
Qy 5101 ATTTGGTAGTGTCTGGGGAGTGGAATCTCCCGGATAACTGCCGTGGTGGTGTCACTGTTTT 5160
Db 5101 ATTTGGTAGTGTCTGGGGAGTGGAATCTCCCGGATAACTGCCGTGGTGGTGTCACTGTTTT 5160
Qy 5161 GTATTGTAGATAAGAAATGAAAGGAGTAGAGGAAGCAACGCTGGGTGCGTATCACGCC 5220
Db 5161 GTATTGTAGATAAGAAATGAAAGGAGTAGAGGAAGCAACGCTGGGTGCGTATCACGCC 5220
Qy 5221 CTGCTTCGAAAAAGAAATTTCTTTTAAAGCTAATCCCTAATTTTCAATAACATCCGAGG 5280
Db 5221 CTGCTTCGAAAAAGAAATTTCTTTTAAAGCTAATCCCTAATTTTCAATAACATCCGAGG 5280
Qy 5281 ATGCTGAGAGACCCCGTGCGCACTGTTAGTGAATATCAAAAGGAGTGGCTATGGAAGAG 5340
Db 5281 ATGCTGAGAGACCCCGTGCGCACTGTTAGTGAATATCAAAAGGAGTGGCTATGGAAGAG 5340
Qy 5341 GATAGTCTCTTTATCTTTGGAGTTCGTTTTCAATTTGTGTAGTACATAAAAAATATGTAA 5400
Db 5341 GATAGTCTCTTTATCTTTGGAGTTCGTTTTCAATTTGTGTAGTACATAAAAAATATGTAA 5400
Qy 5401 GAAAGGTTTGAGGGAACGTTATTTTGTAGTGTGACAGAGCGGTCCGCCAATTGAACTCACTG 5460
Db 5401 GAAAGGTTTGAGGGAACGTTATTTTGTAGTGTGACAGAGCGGTCCGCCAATTGAACTCACTG 5460
Qy 5461 AAAAGGTTTGTGAGGAGTCTGTGATGAAGTACCAGTGGCTGCAAACTCGAAAAGGTTTC 5520
Db 5461 AAAAGGTTTGTGAGGAGTCTGTGATGAAGTACCAGTGGCTGCAAACTCGAAAAGGTTTC 5520
Qy 5521 CGGAAAAACAAAAAGAAATGCTAGTAAATATGTTAATAATAAGAAATAAATACAGTG 5580
Db 5521 CGGAAAAACAAAAAGAAATGCTAGTAAATATGTTAATAATAAGAAATAAATACAGTG 5580
Qy 5581 GTAAGAGGGTTTTAAATTCAGGAATTTAGGATAATGTTAAGTGTGATGACAGTCTATCG 5640
Db 5581 GTAAGAGGGTTTTAAATTCAGGAATTTAGGATAATGTTAAGTGTGATGACAGTCTATCG 5640
Qy 5641 CGTCATCGAGTACGTTTTTAATCAATATGCTTTATACATCAACTCTCCGACCAATTTGT 5700
Db 5641 CGTCATCGAGTACGTTTTTAATCAATATGCTTTATACATCAACTCTCCGACCAATTTGT 5700
Qy 5701 TTACTTATCTTCCGCTTACGAGATCTCTGTGCAGCTGATCAATCTGTGTACAAATGCATT 5760
Db 5701 TTACTTATCTTCCGCTTACGAGATCTCTGTGCAGCTGATCAATCTGTGTACAAATGCATT 5760
Qy 5761 GGGTAAACAGTTTCAACGCAACAGCTAGGACAAACAGTCCAAACAGCAATTTGCGGATGC 5820
Db 5761 GGGTAAACAGTTTCAACGCAACAGCTAGGACAAACAGTCCAAACAGCAATTTGCGGATGC 5820
Qy 5821 CTGGAACCTGTGCTAGTATGACAGTGTGAGATTTCTCGCATCGGATTTCTATGTGTATAG 5880
Db 5821 CTGGAACCTGTGCTAGTATGACAGTGTGAGATTTCTCGCATCGGATTTCTATGTGTATAG 5880
Qy 5881 ATATAATTCGACGCTTGTATCGGCTTGTATCAAGGCTTATTAATAGCTTTGATACAGAA 5940
Db 5881 ATATAATTCGACGCTTGTATCGGCTTGTATCAAGGCTTATTAATAGCTTTGATACAGAA 5940
```

```
Qy 5941 TAGAATAATAGAGTGTGATAATCAACCCGACCGAATACTACTGAAATCGTTAACGGAC 6000
Db 5941 TAGAATAATAGAGTGTGATAATCAACCCGACCGAATACTACTGAAATCGTTAACGGAC 6000
Qy 6001 TCAGAGGGTAGACGATGCTACTGTAGCTATATAAGGGCTTCAATCAATAATTTGGCTAATGA 6060
Db 6001 TCAGAGGGTAGACGATGCTACTGTAGCTATATAAGGGCTTCAATCAATAATTTGGCTAATGA 6060
Qy 6061 ACTGGTTTCGTGGAACTGGGCATGTTCAATCAACGAGGCTTTGAGACTGCTAGTGACATTGT 6120
Db 6061 ACTGGTTTCGTGGAACTGGGCATGTTCAATCAACGAGGCTTTGAGACTGCTAGTGACATTGT 6120
Qy 6121 CTGACACCAACTCCGGCTACTTAGCTATTGTTGTGAGATTTCCTAAAAATAAAGTCGCTG 6180
Db 6121 CTGACACCAACTCCGGCTACTTAGCTATTGTTGTGAGATTTCCTAAAAATAAAGTCGCTG 6180
Qy 6181 AAGACTTTAAAAATTCAGGGTGGCTGATACCAAAATCAGCAGTGGTTGTTCCGCTCACTTAAA 6240
Db 6181 AAGACTTTAAAAATTCAGGGTGGCTGATACCAAAATCAGCAGTGGTTGTTCCGCTCACTTAAA 6240
Qy 6241 TATAACGATTGTCTATCTCGATCCACAGTTAAACCATGTGATGGTGTATCTACTGTGTA 6300
Db 6241 TATAACGATTGTCTATCTCGATCCACAGTTAAACCATGTGATGGTGTATCTACTGTGTA 6300
Qy 6301 TGGCGTAAACATCGGAGAGTTCGAATCTCCCTAACCCGCGGTAGCGGCCCA 6355
Db 6301 TGGCGTAAACATCGGAGAGTTCGAATCTCCCTAACCCGCGGTAGCGGCCCA 6355
```

## RESULT 4

```
AB078435
LOCUS AB078435 6356 bp RNA linear VRL 22-JAN-2002
DEFINITION Tobacco mild green mosaic virus complete genome, strain:Japanese.
ACCESSION AB078435
VERSION AB078435.1 GI:18253266
KEYWORDS
SOURCE Tobacco mild green mosaic virus
ORGANISM Tobacco mild green mosaic virus
REFERENCE 1 Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
AUTHORS Okuno, T., Hamada, H., Takeuchi, S., Morishima, N., Yoshimoto, E. and Hikichi, Y.
TITLE Nucleotide sequence of the Japanese isolate of Tobacco mild green mosaic virus
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 6356)
AUTHORS Okuno, T., Hamada, H., Takeuchi, S., Morishima, N., Yoshimoto, E. and Hikichi, Y.
TITLE Direct Submission
JOURNAL Submitted (18-JAN-2002) Tetsuro Okuno, Kyoto University; Kitashirakawa Oiwakecho, Sakyo 606-8502, Japan
(S-mail:okuno@kais.kyoto-u.ac.jp. Tel:81-75-753-6131, Fax:81-75-753-6131)
```

## FEATURES source

```
1..6356
/organism="Tobacco mild green mosaic virus"
/virus
/mol_type="genomic RNA"
/strain="Japanese"
/db_xref="taxon:12241"
/notes="RNA transcripts synthesized in vitro from the cDNA clone of TMGMV-J direct systemic infection in Nicotiana benthamiana-TMGV-Japanese strain"
72..4901
/notes="position 3405..3407 is readthrough product"
```

## CDS

```
/codon_start=1
/transl_except=(pos:3405..3407,aa:OTHER)
/product="183kDa protein"
/protein_id="BAB93986.1"
/db_xref="GI:18253267"
/translation="MAHIQSTISNALLSVSGKNTLVNDLARRRMYDTAVDFNARDR
RPKNPPTTISEQTLVATYEPQITFYNTONAVHSLAGLPALEVLMLQPVG
SPYTDIGNFAHLFKGRDVFHCMPNLDIRDINRHEGQKDSIEMYLRLSRNKPVP
```





Qy	1082	GACGCTTTGCTTACAAGAAACCTTTGGCCATGTTTCAACACTGAAGAGCAATCTTTAGA	1141
Db	1083	GACGCTTTTACTTATAAGAAAGCCCTTGCTATGTTTCAACACTGAAGAGCAATCTTAA	1142
Qy	1142	GACACGGCTTCGGTTAACTTTTGGTTCCCTAAGATGAAGGACATGGTGATGATGACCGCTG	1201
Db	1143	GACACGGCTTCGGTTAATTTTGGTTTCCCAAGATGAAGGACATGGTGATGATGACCGCTG	1202
Qy	1202	TTTGAGGGTTCTATTATCCAGCAAAAAGATGA CAAGGAGTGAGGTCATTTGTTTAATCGTGAC	1261
Db	1203	TTTGAGGGTTCTATTATCCAGCAAAAAGATGA CAAGGAGTGAGGTCATTTGTTTAATCGTGAC	1262
Qy	1262	TTTGCTTTACACAGTGCTTAAATCATATCAGAACATATCAAGCCAAAGCGTTAACTTACCAG	1321
Db	1263	TTTGCTTTATACAGTGCTTAAATCATATCAGGACATATCAAGCCAAAGCGTTAACTTACCAG	1322
Qy	1322	AACGTTATATCTTTTCGTTGGAGTCTATAAGATCCCGCGTGATAATCAATGGTGTTTACGTCT	1381
Db	1323	AACGTTATATCTTTTCGTTGGAAATCTATAAGATCCCGCGTGATAATCAATGGTGTTTACGTCT	1382
Qy	1382	AGGTCCTGAATGGGATGTAGATAAAGCAATCTTCAACCCCTTGTCAATGAAGCTTTTCTCTTG	1441
Db	1383	AGGTCCTGAATGGGATGTAGATAAAGCAATCTTCAACCCCTTGTCAATGAAGCTTTTCTCTTG	1442
Qy	1442	CAGACTAAGCTGGCTGGCTTCAAGACGATATAGTAATGGGAAAGTTTCGGTGCTTGGAT	1501
Db	1443	CAGACTAAGCTGGCTGGCTTCAAGATGATATAGTAATGGGAAAGTTTCGGTGCTTGGAT	1502
Qy	1502	AAGACCACTTCTGAACCTATTTTGGGATGAGTGCGGCAAAATTTTGTGAAGAGTTTCCCC	1561
Db	1503	AAGACCACTTCTGAACCTATTTTGGGATGAGTGCGGCAAAATTTTGTGAAGAGTTTCCCC	1562
Qy	1562	ACTATCAAAGAGAGATTTGGTGAGCAGGAAATTTCTCGATGTAACTGAGAAATGCTCTGAAG	1621
Db	1563	ACTATCAAAGAGAGATTTGGTGAGCAGGAAATTTTGGATGTAGTGAGTGAATGCTCTGAAG	1622
Qy	1622	ATCAAGATCCAGATCTGTATGTGCATATGGAAGACAGGTTCTGTAGCTGAATACCAAG	1681
Db	1623	ATCAAGATCCCGATCTGTATGTCACTGGGAAGACAGGTTCTGTAGCTGAATACCAAG	1682
Qy	1682	TCTGAGAGTTACCGCATCTAGATATCAAGAGGACTTTAGAGAGCTGAGCAAAATGTAC	1741
Db	1683	TCTGAGAGTTACCGCATCTGATATCAAGAGGACTTTAGAGAGCTGAGCAAAATGTAC	1742
Qy	1742	GACCGGTTATCAGAAATATCTATCTTTAAGGGTCTGTATATTTTGGATATCGCGAAGTTC	1801
Db	1743	GACCGGTTATCAGAAATATCTATCTTTAAGGGATGTGTATTTTGGATATCGCGAAGTTC	1802
Qy	1802	AAAGACATGTCAAGGCTTTAGATGTTAGTCTCGATGTGGCAGCA CGAGTAATCGTTGCA	1861
Db	1803	AAAGACATGTCAAGGCTTTATGTGTTAAATCTCTGATGTGGCAGCA CGAGTAATCGTTGCA	1862
Qy	1862	GTGSCCGAGATAGAAGCGGTTTAACTTTACTTTTGTATAGCCCAACCGAGGAGAAATGTG	1921
Db	1863	GTGSCCGAGAAATAGAAGCGGTTTGAATCTTTACTTTTGTATAGCCCAACCGAGGAGAAATGTG	1922
Qy	1922	GCTAAGGCTCTTAAAGCAGCGGCTCTGAGGCGCTGTGTGTCTTTGAACCGACATCGGAA	1981
Db	1923	GCTAAGGCTCTTAAAGCAGCGGCTCTGAGGCGCTGTGTGTCTTTGAACCGACATCTGAT	1982
Qy	1982	GAGGTGAACGTAAATAAATTTTCTATTGCTGAGAAAGGAGATGTGCTGTGTGTCAGAA	2041
Db	1983	GAGGTGATGTAATAAATAAATTTTCTATTGCTGAGAGGGGAAATTTGCGGCTGTGTGCTGAA	2042
Qy	2042	AGTCATGGTTGACGAATGCTAACTTTAGAGCACGAGGTTGGAGTCCCTCAACGATTTTC	2101
Db	2043	AGTCATGGTTGACGAATGCTAACTTTAGAGCACGAGGTTGGAGTCTCTCAACGATTTTC	2102
Qy	2102	CATAAGGCTTCGCTGGATAGTGATTTTACAAAGCAAATGGCATCGGTTGTCTACACTGGC	2161
Db	2103	CATAAGGCTTCGCTGGATAGTGATTTTACAAAGCAAATGGCATCGGTTGTCTATCTATCTGGC	2162
Qy	2162	TCACCTCAAAGTTCAACAAATGAAGAACTATGTGGAAGTTTGGCAGCTTCGTTGTTCGCGC	2221

2163	TCACTCAAAGTTCCAAACAATGAAGAACTAATGTGGACAGTTCGCGAGCTTCGTTGTCGCGC	2222
2222	ACTGTATCAAATCTATGCGAAGTCACATAAAGGATGAAGTCGGGTATGATTCTGATTTCCAGG	2281
2223	ACTGTATCAAATCTATGCGAAGTCATTAAGAAGATGTTGTTGGGTATGATTTCTGATTTCCAGG	2282
2282	GAGAAAGTTGGTGTGTTGGGATGTCACTTTTGAAAAAGTGCGTCTCTCAAACCTCGGCGCAAA	2341
2283	GAGAAAGTTGGTGTGTTGGGATGTCACTTTTGAAAAAGTGCGTCTCTTTAAACCGGCGGC AAAA	2342
2342	GGTCATTTCATGGGGAGTTGTCCTGGATTAACAAGGGGAAAATGTTTACTGCACTTCTATCT	2401
2343	GGCCATTTCAATGGGGAGTTGTCCTGGATTAACAAGGGGAAAATGTTTACC GCATCTTCTATCT	2402
2402	TATGAAGGAGATAGAATGGTGACTGACAGCGACTGGAGGAGCGTGGCTGTATCATCTGAT	2461
2403	TATGAAGGAGATAGAATGGTGACTGACAGCGACTGGAGGAGCGTGGCTGTCTCATCTGAT	2462
2462	ACAATGGTATATTCTGATATTGCCAATGTCCAAAGCTCCAAAATCTCGAGAAAACAATAGAGACGGT	2521
2463	ACGATGGTATATTCTGATATTGCGAAGCTCCAAAATCTCGAGAAAACAATAGAGACGGT	2522
2522	GAACCCCA CGAACCTTACTGCA AAGATGGTACTTGTGGATGGGGTGGCTGGTTGTGGAAAG	2581
2523	GAACCCCA CGAACCTTACTGCA AAGATGGTACTTGTGGATGGGGTGGCTGGTTGTGGAAAG	2582
2582	TACAAAGGAGATTTTTCA AAGATTTGATCTTTGATGAGGATTTTGATCTTTGGTTCCTCGMAAA	2641
2583	ACAAGAGGATTTTTGAAAGAGTTGATCTTTGATGAGGATTTGATCTTTGGTTCCTGGMAAA	2642
2642	C AAGCTGCTATGATCAGAAAGGGCTAAATTCATCTGGACTGATTAAGAGCCCAATG	2701
2643	C AAGCTGCTATGATCAGAAAGGGCTAAATTCATCTGGACTGATTAAGAGCCCAATG	2702
2702	GACAATGTGAGAACGGTAGATTCACTTTAATGCACTCCAAAACCGGATCA CACAAGAGG	2761
2703	GACAATGTGAGAACGGTAGATTCACTTTAATGCACTCCAAAACCGGATCA CACAAGAGG	2762
2762	CTTTTAAATGATGAAGGGTTGATGTCACACCGGTTGTGTTAACTTCTCGTGTCTTATC	2821
2763	CTTTTCAATGATGAAGGGTTGATGTCGACACCGGTTGTGTTAACTTCTCGTGTCTTATC	2822
2822	TCTGGTTGGCACATCGCATACATTTACGGAGATACACAGCAGATTCCTTTCAATTAACAGA	2881
2823	TCTGGTTGGCACATCGCATACATTTACGGAGATACACAGCAGATTCCTTTCAATTAACAGA	2882
2882	GTTTCAGAAATTCCTGATATCCC AAA CAATTTTGAGAGCTGCAAGTGGATGAAGTTGAGATG	2941
2883	GTTTCAGAAATTCCTGATATCCC AAA CAATTTTGAGAGCTGCAAGTGGATGAAGTTGAGATG	2942
2942	AGGAGACCA CATGAGATGCCAGGTGATGAAATTTTTTCTCAATCGAAGTACGAA	3001
2943	AGGAGACCA CATGAGGTGCCAGGTGATGAAATTTTTTCTCAATCGAANAACGAA	3002
3002	GGAGCGGTGCAACCACTTCAACTGTACAAAGATCGGTCTCATCTGAGATGATAGCGGT	3061
3003	GGAGCGGTGCAACCACTTCAACTGTACAAAGATCAGTCTCATCTGAGATGATAGAGGT	3062
3062	AAGGGAGTACTAAACAGTGTTC AAA CCACTTAAAGGGAAAATTTGTAACTTTCACTGAG	3121
3063	AAGGGAGTACTAAACAGTGTTC AAA CCACTTAAAGGGAAAATTTGTAACTTTCACTGAG	3122
3122	GCTGATAAATTTGAGTTAGAGGAGGGGCTAAGAAATGTC AACACCGTTCATGAGATC	3181
3123	GCTGATAAATTTGAGTTAGAGGAGGGGCTAAGAAATGTC AACACCGGTTCA TGAGATC	3182
3182	CAAGGAGAAACCTTTGAGATGTGCGTGGTCAGATTTGACGGCAACTCCA CTGACTCTG	3241
3183	CAAGGAGAAACCTTTGAGATGTGCGTGGTCAGATTTGACAGCAACTCCA CTGACTCTG	3242
3242	ATTTCCAGGCTTTCCCGCATGTTCTGATGCGTCTGACTGACACACAAGAGCTTTCAA	3301

Db 3243 ATTTCCAAAGTCTTCCCGCATGTTCTTAGTCGCGCTGACTAGACACACAAAGAGCTTCAAA 3302  
Qy 3302 TATTACACCGTAGTGTAGATCTCTTTAGTACAGATAAATAGTAGATTGTCTTCTTTAAAGC 3361  
Db 3303 TATTACACCGTAGTGTAGATCTCTTTAGTACAGATAAATAGTAGATTGTCTTCTTTAAAGC 3362  
Qy 3362 TCCCTTCCCTTTAGAAATCTATATGTAGAACGAGTAGTAGATAGCAATTAACAGTGGAT 3421  
Db 3363 TCCCTTCCCTTTAGAAATCTATATGTAGAACGAGTAGTAGATAGCAATTAACAGTGGAT 3422  
Qy 3422 GCAGTGTTCAAAAGGTCATAATCTCTTTGTGGCAACACCTAAATCAGGAGACTTTCCAGAT 3481  
Db 3423 GCAGTGTTCAAAAGGTCATAATCTCTTTGTGGCAACACCTAAATCAGGAGACTTTCCAGAT 3482  
Qy 3482 CTACAGTGTCTAATAGATAGTATGCTCCCTGGTAAATAGTACTATATCTTAAACAAGTAGAT 3541  
Db 3483 CTGAGTGTCTAATAGTATGATGTCTCCCTGGTAAATAGTACTATATCTTAAACAAGTAGAT 3542  
Qy 3542 GCTGTATACCATGAGGTTACGGGTAATAGTCTTAAATGTGAAGGATTTGTGTTCTTGAATTT 3601  
Db 3543 GCTGTATACCATGAGGTTACGGGTAATAGTCTTAAATGTGAAGGATTTGTGTTCTTGAATTT 3602  
Qy 3602 TCCAAAAGTATTCGGATCGCAAGAGGTGAACCATGTCTAGAGCCAGTTTTCGCTACC 3661  
Db 3603 TCCAAAAGTATTCGGATCGCAAGAGGTGAACCATGTCTAGAGCCAGTTTTCGCTACC 3662  
Qy 3662 GCGCGGAACCGCAAGGCTGCAAGGACTCTCGAAATCTGGTTGCAATGATTAAGA 3721  
Db 3663 GCGCGGAACCGCAAGGCTGCAAGGACTCTCGAAATCTGGTTGCAATGATTAAGA 3722  
Qy 3722 AATTTCACGACACAGACCTGACGGGACGATTCACATTTAGAGACACCGCATCTTTGTA 3781  
Db 3723 AATTTCACGACACAGACCTGACGGGACGATTCACATTTAGAGACACCGCATCTTTGTA 3782  
Qy 3782 GTAGATAAGTTTTTGTAGTACTATTTTATTAAGAAAGAAATACACAAAATAATTGCT 3841  
Db 3783 GTAGATAAGTTTTTGTAGTACTATTTTATTAAGAAAGAAATACACAAAATAATTGCT 3842  
Qy 3842 GGAAGTAGACGAAGGATTCATGATGAGATGGTTTGGGAAACAGGAGAGTACTATC 3900  
Db 3843 GGAAGTAGACGAAGGATTCATGATGAGATGGTTTGGGAAACAGGAGAGTACTATC 3902  
Qy 3901 GGACGACTTGCTTAACATAATTTTACAGATCTCGCGGCATCGATCAGTACAGCACAT 3960  
Db 3903 GGAC - AGTTGGCTTAACATAATTTTGTAGATCTCGCGGCATCGACAGTATAAGCACAT 3961  
Qy 3961 GATCAAGGCTCAACCAAAACAGAAATTTGACCTTTCAATTCAGAAATGAATACCCGTCT 4020  
Db 3962 GATCAAGGCTCAACCAAAACAGAAATTTGACCTTTCAATTCAGAAATGAATACCCGTCT 4021  
Qy 4021 GCAAAACAAATTTCTACCAATTCGAAGCAGATCAACCGGTATTTTGGCCCGGTTTTTTCAGA 4077  
Db 4022 GCAAAACAAATTTCTACCAATTCGAAGCAGATCAACCGGTATTTTGGCCCGGTTTTTTCAGA 4081  
Qy 4078 GCTTACAGGTTGCTGCTCGAGCATTTGATCTTAAGAGTTCTTTTCTTACTAGGAA 4137  
Db 4082 GCTTACAGGTTGCTGCTCGAGCATTTGATCTTAAGAGTTCTTTTCTTACTAGGAA 4141  
Qy 4138 AACTCCAGAACAGATTCAGAAATTTTCTCGATCTCGACTCGACGCTTCTATGGATGT 4197  
Db 4142 AACTCCAGAACAGATTCAGAAATTTTCTCGATCTCGACTCGACGCTTCTATGGAGCT 4201  
Qy 4198 GTTAGAACTGGATATTTCTAAGTAGATGAATGTCACAGAACGAGTTTCATTGTGCTGAGA 4257  
Db 4202 GTTAGAACTGGATATTTCTAAGTAGATGAATGTCACAGAACGAGTTTCATTGTGCTGAGA 4261  
Qy 4258 GTATGAATATGGAAAGATTTGGTCTCAATGAGTTTGGCCGAGTGTGGAAACAAGG 4317  
Db 4262 GTATGAATATGGAAAGATTTGGTCTCAATGAGTTTGGCCGAGTGTGGAAACAAGG 4321  
Qy 4318 GCACAGAAACAACTTTTGAAGGATTAATGCTGGAAATCAAGACATGCTGTGGTATCA 4377  
Db 4322 ACACAGGAAACAACTTTTGAAGGATTAATGCTGGAAATCAAGACATGCTGTGGTATCA 4381

Qy 4378 AAGGAAAGCGGTGATGTGACTACTTTTCATCGGCAATACTGTATATAATAGCAGCTTGCTT 4437  
Db 4382 AAGGAAAGCGGTGATGTGACTACTTTTCATCGGCAATACTGTATATAATAGCAGCTTGCTT 4441  
Qy 4438 GGGTTCAATGTTACCGATGGAAAGGTCATAAAAGTGCTTTTGTGGAGACGATCCGT 4497  
Db 4442 GGGCTCAATGTTACCTATGGAAAGGTCATAAAAGTGCCCTTTTGTGGAGATGACTCGT 4501  
Qy 4498 TTTGTATTTTCCAAAGGTTTGGATTTCCCTGACATTCAGTCATGTCTAATCTCATGTG 4557  
Db 4502 TTTGTATTTCCGGAAGGTTTGGATTTCCCTGACATTCAGTCATGTGCCATCTCATGTG 4561  
Qy 4558 GAATTTTGAAGCCAAACTGTANTAGAAAGAGTACGGTTACTTTTGTGGTAGATACATCAT 4617  
Db 4562 GAATTTTGAAGCCAAACTGTACAGAAAGAGTACGGTTACTTTTGTGGTAGATACATCAT 4621  
Qy 4618 ACACCATGATAAGGGAGCAATAGTATTAATGATCTTGAAGTGTGATCTCCAACTGG 4677  
Db 4622 ACACCATGATAAGGGAGCAATAGTATTAATGATCTTGAAGTGTGATCTCCAACTGG 4681  
Qy 4678 GGCATAAATCATATCAAGGATTAATGATCACTTAGAAGAGTTAAGGCTGCTTTGTGCGATGT 4737  
Db 4682 GGCATAAATCATATCAAGGATTAATGATCACTTAGAAGAGTTAAGGCTGCTTTGTGCGATGT 4741  
Qy 4738 TGTGTGTTGCTCGGAAACTGTGCTTTAGGCTTTCCGACGCTGAACCGACGCTATCAAGGA 4797  
Db 4742 TGTGTGTTGCTCGGAAACTGTGCTTTAGGCTTTCCGACGCTGAACCGACGCTATCAAGGA 4798  
Qy 4798 GGTTCATAAACCAGGATTCATGCTGCTTTGCTTTTAAATGTTGTTAAACAAATTTTGTG 4857  
Db 4799 GGTTCATAAACCAGGATTCATGCTGCTTTGCTTTTAAATGTTGTTAAACAAATTTTGTG 4858  
Qy 4858 TGATAAATTTTATTTAGAACTTTGTTTAAATGCTGTTAGTCTCAGAGATACTGTCA 4917  
Db 4859 TGATAAATTTTATTTAGAACTTTGTTTAAATGCTGTTAGTCTCAGAGATACTGTCA 4918  
Qy 4918 AATTTAGCGAGTTCATTTGATCTTTGAAACAGGATGAGATCTTCCGGCATTCATGACTA 4977  
Db 4919 AATTTAGCGAGTTCATTTGATCTTTGAAACAGGATGAGATCTTCCGGCATTCATGACTA 4978  
Qy 4978 AGTCAAGAGTGTAGAAATCGACTGTGACAGATTAATGCTGTTTAAAGAAAGGATGATGTC 5037  
Db 4979 AGTCAAGAGTGTAGAAATCGACTGTGACAGATTAATGCTGTTTAAAGAAAGGATGATGTC 5038  
Qy 5038 TTTCTGATGTAGATTTTACTTTAAAGGTTTAAAGTGTAGTAAAGAAAGGATGATGTCGTTAG 5097  
Db 5039 TTTCTGATGTAGATTTTACTTTAAAGGTTTAAAGTGTAGTAAAGAAAGGATGATGTCGTTAG 5098  
Qy 5098 CTGATTTGTTAGTGTCTGGGAGTGGAAATCTCCCGGATACTGCGGTTGGTGTCAAGT 5157  
Db 5099 CTGATTTGTTAGTGTCTGGGAGTGGAAATCTCCCGGATACTGCGGTTGGTGTCAAGT 5158  
Qy 5158 TTTGTATTTGATAGTAAGAAATGAAAGGAGTAAAGGAGCAACGCTGGGTTGATCAAG 5217  
Db 5159 TTTGTATTTGATAGTAAGAAATGAAAGGAGTAAAGGAGCAACGCTGGGTTGATCAAG 5218  
Qy 5218 CCCTGCTTCCAAAAGAAATTTTCTTTAAAGCTTAATCCCTAATTTTCAATAACATCCG 5277  
Db 5219 CCCTGCTTCCAAAAGAAATTTTCTTTAAAGCTTAATCCCTAATTTTCAATAACATCCG 5278  
Qy 5278 AGGATGCTGAGAACGCCCGGTGGCAAGTGTGTAGTAATCAAGAGGAGTGGCTATGGAAG 5337  
Db 5279 AGGATGCTGAGAACGCCCGGTGGCAAGTGTGTAGTAATCAAGAGGAGTGGCTATGGAAG 5338  
Qy 5338 AAGGATGCTGCTTTTATCTTTGAGTTCGTTTCAATTTTGTGTAGTACATAAATAAATG 5397  
Db 5339 AAGGATGCTGCTTTTATCTTTGAGTTCGTTTCAATTTTGTGTAGTACATAAATAAATG 5398  
Qy 5398 TAAGAAAGGTTTGAAGGAACTGATTTTGTAGTGTGACAGACGCTCGCCAAATGAACCTCA 5457  
Db 5399 TAAGAAAGGTTTGAAGGAACTGATTTTGTAGTGTGACAGACGCTCGCCAAATGAACCTCA 5458

QY 5458 CTGAAAGGTTGTTGAGGAGTTCGTGGATGAAGTACCAATGGCTGTGAACCTCGAAAGG 5517  
Db CTGAAAGGTTGTTGAGGAGTTCGTGGATGAAGTACCAATGGCTGTGAACCTCGAAAGG 5518  
QY 5518 TTCGGGAAACAAAAAGAAATGTTAGTAAATAATGTTAAATAAAGAAATAAATAACA 5577  
Db 5519 TCCGGGAAACAAAAAGAGAGTGTGTAGTAAATAATGTTAAATAAAGAAATAAATAA 5578  
QY 5578 GTGGTAGAGAGGTTTAAATTTGAGGAAATGAGGATTAATGTAAGTGTACAGAGCTTA 5637  
Db 5579 GTGGTAGAGAGGTTTAAATTTGAGGAAATGAGGATTAATGTAAGTGTACAGAGCTTA 5638  
QY 5638 TCAGGTCAATCAGTACGTTTAAATCAATATCCCTTATACCAATCAACTCTCCGAGCCAAAT 5697  
Db 5639 TCAGGTCAATCAGTACGTTTAAATCAATATCCCTTATACCAATCAACTCTCCGAGCCAAAT 5698  
QY 5698 TGTTTACTTATCTTCCGCTTACGAGATCTCTGTGAGTCAATCTGTGTACAAATGC 5757  
Db 5699 TGTTTACTTATCTTCCGCTTATGAGATCTCTGTGAGTCAATCTGTGTACAAATGC 5758  
QY 5758 ATTGGGTAAACAGTTTCAAGGAAACAGCTAGGACCAACAGTCCAAAGCAATTTGGCGGA 5817  
Db 5759 ATTGGGTAAACAGTTTCAAGGAAACAGCTAGGACCAACAGTCCAAAGCAATTTGGCGGA 5818  
QY 5818 TGCTCTGGAACCTGTGCTAGTATGACAGTCAAGTATTCCTGATCGGATTTCTATGTGTA 5877  
Db 5819 TGCTCTGGAACCTGTGCTAGTATGACAGTCAAGTATTCCTGATCGGATTTCTATGTGTA 5878  
QY 5878 TAGATATAATTCGAGCTTGTATCGGTTGATCAACCGGCTTATTAATAGCTTTGATAGTAG 5937  
Db 5879 TAGATATAATTCGAGCTTGTATCGGTTGATCAACCGGCTTATTAATAGCTTTGATAGTAG 5938  
QY 5938 AAATAGATAATAGAGGTTGATATCAACCGGACCGAATCTAATCTGAATTCGTTAAACGC 5997  
Db 5939 AAATAGATAATAGAGGTTGATATCAACCGGACCGAATCTAATCTGAATTCGTTAAACGC 5998  
QY 5998 GACTCAGAGGTAGACAGTCTAGTCTAGTATGAGGCTTCAATCAATTAATTTGGCTAA 6057  
Db 5999 GACTCAGAGGTAGACAGTCTAGTCTAGTATGAGGCTTCAATCAATTAATTTGGCTAA 6058  
QY 6058 TGAATCGTTTCTGGAACCTGCGATGTTCAATCAAGCAGGCTTTGAGAGTCTGCTAGTACT 6117  
Db 6059 TGAATCGTTTCTGGAACCTGCGATGTTCAATCAAGCAGGCTTTGAGAGTCTGCTAGTACT 6118  
QY 6118 TGCTCGACCAAACTCCGGTACTTATGAGTATTTGAGATTTCTTAAATAAAGTCG 6177  
Db 6119 TGTTTGCACCAAACTCCAGTACTTATGAGTATTTGAGATTTCTTAAATAAAGTCG 6178  
QY 6178 CTGAGACTTAAATTCAGGCTGCTGATACCAAAATCAGCAGTGGTTGTCGCTCACTT 6237  
Db 6179 CTGAGACTTAAATTCAGGCTGCTGATACCAAAATCAGCAGTGGTTGTCGCTCACTT 6238  
QY 6238 AAATATACAGTCTCATATCTGGATCAACAGTTTAAACCAATGATGTTGATGATGCTG 6297  
Db 6239 AAATATACAGTCTCATATCTGGATCAACAGTTTAAACCAATGATGTTGATGATGCTG 6298  
QY 6298 GTATGGGTAAACATCGGAGGTTGGAATCTTCCCTTAAACCGGCTAGCGGCCCA 6355  
Db 6299 GTATGGGTAAACATCGGAGGTTGGAATCTTCCCTTAAACCGGCTAGCGGCCCA 6356

## RESULT 5

E07987  
LOCUS E07987  
DEFINITION Tobacco mosaic virus L cDNA.  
ACCESSION E07987  
VERSION E07987.1 GI:2176118  
KEYWORDS JP 1994233688-A/1.  
SOURCE unidentified  
ORGANISM unclassified  
REFERENCE 1 (bases 1 to 6384)  
AUTHORS Murai,F., Akiyama,T., Watanabe,Y. and Okada,Y.

TITLE  
JOURNAL

TRANSCRIPTION VECTOR FOR IMPROVING PLANT VIRUS  
Patent: JP 1994233688-A 1 23-AUG-1994;  
SANKYO CO LTD  
OS (Tobacco mosaic virus L)  
PN JP 1994233688-A/1  
PD 23-AUG-1994  
PF 29-NOV-1993 JP 1993297704  
PR 01-DEC-1992 JP 92P 322100  
PI MURAI FUKASHI, AKIYAMA TOSHIYUKI, WATANABE YUICHIRO, PI  
OKADA YOSHIMI

## COMMENT

CC strandedness: Double;  
CC topology: Linear;  
CC Feature is identified by experimental;  
FH Key Location/Qualifiers

FT source 1..6384  
FT FT /organism='Unclassified'  
FT 5'UTR 1..71  
FT CDS 72..3422  
FT /product='130kd-replication enzyme' FT CDS  
FT 72..4919  
FT /product=(seq:'tag', aa:Tyr)  
FT CDS 4906..5700  
FT /product='movement protein'  
FT CDS 5703..6182  
FT /product='coat protein'  
FT 3'UTR 6183..6384.  
FEATURES  
source  
1..6384  
/organism="unidentified"  
/mol\_type="genomic RNA"  
/db\_xref="taxon:32644"

## ORIGIN

Query Match 41.4%; Score 2631; DB 6; Length 6384;  
Best Local Similarity 64.2%; Pred. No. 0;  
Matches 4093; Conservative 0; Mismatches 2230; Indels 54; Gaps 7;  
QY 18 CGACAAACAATTAACAAACAAACATATTACAAACAAACAAACAAACATGCGAC 77  
Db 23 CAACACACAAACAAACAAACATATTACATTTTACATTC---TACAACTACATGCGAT 78  
QY 78 ACATACATCTAATAATTAGCAACGCCCTTCTTGAAGCGTGAGTGGTAAACAACTCTCG 137  
Db 79 ACACAAACAGCACATCGTCGCTTCTTGAGACCGTCCGAGGTAAACATACCTTGG 138  
QY 138 TTAATGACCTTGAAGAGCGCGATGTACGATACGCGCGTGGGAAGAAATTTAAACCCCGCG 197  
Db 139 TCAACGATCTTGCANAAGCGCGCTATATGACACAGCGGTGATGAATTTAATGCTAGGG 198  
QY 198 ACCGTAGACCAAGGTTCAACTTTTCCAAACTATTAGCGAAGAGCAACGCTTCTAGTCT 257  
Db 199 ACCGAGGCTTAAAGTCAATTTTCCAAAGTAGTAAGCGAAGAACAGACGCTTATTGCAA 258  
QY 258 CCAACGCGTACCGGAGTTCAGATTAACCTTTTATAATACTCAAAATSCCGGTACACAGTT 317  
Db 259 CCAAGCTTACCAGAAATTCGAATTTACATTTTACAACACGAGAAATGCTGTGCAATCCC 318  
QY 318 TGGCTGAGGTTTGAAGCAATTAAGATTTGAATATCTGATGCTACAAGTTCCCTTATGGAT 377  
Db 319 TTGAGCGGCTCTCCGATCAATTAAGAAATTTGAATATCTGATGATGCAAAATTTCCCTTACGGAT 378  
QY 378 CGCGCATATGATATAGTGGGAATTTTCACACATTTTGTTCAAAGGCGAGGATTAACG 437  
Db 379 CATGACATATGATATCGGAGGTAATTTTGCATCTCATCTGTTCAAAGGGGAGCATACG 438  
QY 438 TGCATTGCTGTATGCCCAATCTCGACATACGAGATATAATGAGGCACGAAGGACAAAGG 497  
Db 439 TTCACTGCTATGCCGAACTCGATGCTCCGACATATGCGCACGAATGCGGACGAGGCCAAGG 498  
QY 498 ACTCAATTGAGATGATATTTGTCAGATTTGTCCTTAAACAGGTAATTTCTCTGAGTTTC 557

Db 499 ACAGTATTGAACTATACCTTTCTAGGCTCGAGAGGGGCAACAATGTCCTCCAACTTCC 558  
Qy 558 AAAGGGAGGCTTTTAAACAGGTATGAGAGCTCCCAACGAAAGTCTGCTGCTCTTAAACTT 617  
Db 559 AAAAGGAAGCTTTTCACAGATACGCTGAAATGCCAAACGAAGTAGTCTGTCAACGATACTT 618  
Qy 618 TTCAGGATGTCGAATACATCGCCAGAGAGATAGTGTGAGAGATACGCTGTTGCTGTC 677  
Db 619 TCCAAACGTGAGGCATCTCAAGAATGTTTACACGGGAAGAGTGTATGCTATTGCTTTCG 678  
Qy 678 ACAGTTTCTATGATATCTCTGCGATGAGTTTGGAGCTGCGCTTAATATCTAAGAATATAC 737  
Db 679 ATAGTATATAGATATACCTCGCGAGAGTTTCGGCGGCGACTGCTGAGAAAGAGATGTAC 738  
Qy 738 ATGATGTTATGCACTTTCCATTTTGGCAGAGCAATTATTAATAGACAGACGGAGTTA 797  
Db 739 ATGATGTTATGCGCTTTCCACTTTTCCGAGAAATTTACTTCTCGAAGATTCACAGGTCA 798  
Qy 798 CGCTTAATGAATAGGCGCACTTTCAAAAGAGAGGTGATGTTCTTTTCTTTTCTTTCG 857  
Db 799 ACCTCGAGAGATCAATGCTATGTTTCCAAAGAGATGGAGACAGGTGTGACTTTTTCCTTTG 858  
Qy 858 CTGATGAAGTACTTTTAAATTTATAGTCAATAATACAAAATATCTTGCAATATGTAGTTA 917  
Db 859 CATCTGAGAGTACTTTTAAATTTATAGTCAATGTTTCTTAATATTTCTTAAGTATGTTGCA 918  
Qy 918 AATCTTACTTTCTGCTTCTAGTGAATAGTTTACTTTTAAAGGAATTTTACTCTACAGGG 977  
Db 919 AAACCTACTTCCAGCCTCTAATAGAGAGTTTACATGAAGGAGTTTATAGTAACATAGAG 978  
Qy 978 TTAATACTGTTGTTGTTAAATTTTCAAAAGTAGATACCTATATTTCTGTACAAAGAGTTA 1037  
Db 979 TTAATACTGTTGTTGTTAAATTTTCTAGAAATAGTACTTTCTTATTTGTACAAAGGTGTAG 1038  
Qy 1038 GACAGTAGGCTGTGATGATGATCACTCACTTATAGGCGATGGAAGAGCCCTTTGCTTACA 1097  
Db 1039 CGCATAGGCTGTAGATAGTACGACGTTTACAGGCTATGGAAGAGCGCATGGCACTACA 1098  
Qy 1098 AGAAACCTTTGGCCATGTTTCAACACTGAAAGAGCAATCTTTTAGACACGGCTTCGGTTA 1157  
Db 1099 AAAAGACTCTTGGCATGTGCAACAGTGAAGAAATCTTGTAGAGGATTTCTTCAACAGTTA 1158  
Qy 1158 ACTTTTGGTTCCTTAAGATGAGACATGTTGATGATACCGCTGTTTGGAGGTTCTATT 1217  
Db 1159 ATTACTGTTTCCAAAATAGGGATATGGTATAGTTTCCACTATTTGACATATCTCTCG 1218  
Qy 1218 CCAGCAAAAGATGACAAAGAGTGAAGTCACTTGTATTAATCGTGACTTCGTTTACACAGTGC 1277  
Db 1219 AGACTAGTAAAGAAACACGCAAGAGTCTTAGTTTCAAGGACTTTGTTTATACAGTGT 1278  
Qy 1278 TTAATCATATCAGAACATATCAAGCCAAAGCGTTTAACTTACCAAGAGATATATCTTTTCG 1337  
Db 1279 TAAATCATATCTGTACGTACCAAGCCAAAGCGCTTACTTACTTCCAAACGTTATCTTTTCG 1338  
Qy 1338 TGGAGTCTTAAGATCCCGCTGTATCATATGTTGTTTACTGCTAGTCTGAAATGGGATG 1397  
Db 1339 TCGAATCAATTCGTTTCGAGAGTATCATATTAACGGGTTTACTGCTAGGTCTGAGTGGGATG 1398  
Qy 1398 TAGATAAGCAATTTCTCAACCTTGTCAATGACTTTTCTTCTGCGAGACTAAGCTGGCTG 1457  
Db 1399 TCGATAATCATATTATACAGTCTCTTGTGCGATGACGTTCTTCTTCTTACACCAAGCTTGGCG 1458  
Qy 1458 CGCTTCAAGACGATATAGTAATGGGAAAGTTTCGGTGTGCTGGAATGAAGCACTTCTGAAC 1517  
Db 1459 TTCTGAAAGACGATCTTTTGTATAGCAAGTTTGCATTTGGACCAAAAATCTGTCTCAACAC 1518  
Qy 1518 TTATTTGGATGAGGTGGGCAATTTTGTGAAACGTTTTCCTCCACTATCAAGAGAGAT 1577  
Db 1519 ATGTGTTGGATGAGATTTTCCCTAGCTTTCCGCAATGCTTTTCCCATCGATCAAGGAAAGAT 1578  
Qy 1578 TGGTGAGCAGGAAATTTCTGATGTAAGTGAAGATGCTCTGAAGATCAAGATCCAGATC 1637

Db 1579 TGATAAACCCGAAACTGATCAAATTTACGGAGAAATGCGTTAGAGATCAGGGTGCCCGATC 1638  
Qy 1638 TGTATGTACATGAAAGACAGGTTTCGTAGCTGAATACACCAAGTCTGAGAGATTACCGC 1697  
Db 1639 TTTATGTCACTTTCCATGATAGGTTAGTTTCTAGTACAAAATGTCAGTGACATGCGG 1698  
Qy 1698 ATCTAGATATCAGAAGGACTTGAAGAAGCTGAGCAAAATGTACGACGCGTTATCAGAAAT 1757  
Db 1699 TGTAGACATATTAGGAAAAAGATGGAAGAAATGAGGAAATGTACAATGCACTGTGCCGAAC 1758  
Qy 1758 TATCTATCTTAAAGGCTGCTGAATAATTTTCGATATCGCGAAAGTTCAAAGACATGTGCAAGG 1817  
Db 1759 TGTCTGTACTTAAAAATTCAGACAAAGTTCCGATGTGACGTTTTCCTCCAGATGTGCCAAT 1818  
Qy 1818 CTTTATAGTATGTCCTGATGTGGCAGACAGATATTCGTTTCGAGTGGCGGAGAAATAGAA 1877  
Db 1819 CTTTAGAAGTCGATCCAAATGACTGACAGCAAAAGGTAATAGTAGCAAGTTATGAGCAACGAGA 1878  
Qy 1878 GCGGTTTAACTCTTACCTTTTGAATAAGCCAAACGAGGAGAAATGTGGCTTAAAGCTCTTAAA- 1936  
Db 1879 GTGGTCTTACTCTCACGTTTGAACAGCCCAACCGAAGCTAATGTTGCGGTAGCATGTCAAG 1938  
Qy 1937 -----AGCACGGCGCTCGAGCCGCTGTATGCTTTGAAACCGACATCCGAAAGAGGTGAACG 1991  
Db 1939 ATTCTGAAAGGCTTCTGATGGGCGTTGGTAGTTTACCTCAAGAGATGTTTGAGGAACCGT 1998  
Qy 1992 TAAATAAATTTTCTATTTCTGAGAAAGAGAGATGTCCTGTGTGTGTCGAGAAAGTCAATGTT 2051  
Db 1999 CCATAAAGGTTTCGATGGCCGCTGATGATGATTACAAATTTGGCGGATTTATCTGGCAGCGTTC 2058  
Qy 2052 TGACGAATGCTTAATAGACACAGGAGTTGGAGTCCCTCAACGATTTCCATGAAGCTT 2111  
Db 2059 CTGAATCTTATACATCCTAGGAGCGAGAGATGAGTCTCTCGACAGTTCCTATATGGCAA 2118  
Qy 2112 GCGTGAAGTGTGATTTACAAAGCAATGGCATCGGTTGTCTACACTGGCTCACTCAAG 2171  
Db 2119 CAGCTAGTTCGTTAAATTCATAAGCAGATGTTTCGATCGTGTACACGCGCCCTCTTAAAG 2178  
Qy 2172 TTCAAAATTAAGAACTATGTGGAAGATTTGGCAGCTTCTGTTGTCGCCCACTGTATCAA 2231  
Db 2179 TTCAACAAATGAATACTTTATAGACAGCCCTGGTAGCTCTGCTCTGCTGCGGTGCGA 2238  
Qy 2232 ATCTATCAAGTCACTAAAGATGAAGTGGGATGATGATTTCTGATTTCCAGGAGAGAAAGTTG 2291  
Db 2239 ATCTAGTGAAGATCTTAAAGATACACCGCGATTTGACCTTGAACCTCGTCAAAAGTTCG 2298  
Qy 2292 GTGTTTGGGATGTCACTTTTGAATAAGTGGCTCTCAAAACCTCGCGCCAAAGGTCAATTCAT 2351  
Db 2299 GAGTTCTGGATGTTGCTTCGAAAAGGTGGCTAGTTAAACCATCGCAAGAACCATGTCAT 2358  
Qy 2352 GGGGAGTTGCTGGATTAACAAGGGAATAATGTTTACTGCACTTCTATCTTATGAAAGGAG 2411  
Db 2359 GGGGGTGTGTGAGACTCATGCGAGGAAATATACGTCGCAATTAATGAGACACGATGAAT 2418  
Qy 2412 ATAGAATGTGACTGAGAGCGACTGAGAGGGGTGGCTGTATCATCTGTATCAATGGTAT 2471  
Db 2419 TTGGCATTTATCTGCGATTAATGCGGACCGGTGGCTGTGAGTTCCTGAGTGGTAGTAT 2478  
Qy 2472 ATTCTGATATTGCAAAAGCTCCAAAATCTGAGGAAAAAACAATGAGAGACGGTGAACCCCAAG 2531  
Db 2479 ATTCTGATATGGCTTAAACTCAGGACTCTGAGAGATGTTGCTCAAAGATGGAAGAACCAACG 2538  
Qy 2532 AACCTACTGCAAGATGTTACTTGTGATGGGTGCTGTTGTTGTTGGAAGATGACAAAGGAG 2591  
Db 2539 TTAGTTTCAAGAAAGGTGGTTTGTGTGATGGCTTCCAGGGTTCGGGGAAGCAAAAGGAAA 2598  
Qy 2592 ATTTTGAAGATTTGATCTTGTAGGAGATTTGATCTTTGGTTTCTCTGGAATAACAAGCTGCTG 2651  
Db 2599 TTCTTTGAGAGTTTATTTTGAAGAGATCTAATTTCTTCTCCCTGCTGCTCAAGCTGCGC 2658  
Qy 2652 CTATGATCAGAAGAGGCTTAATTTCTGAGCTGATGAAGGCCAACATGAGCAATGTGA 2711  
Db 2659 AGATGATCAGAAGAGAGCTAATGCTCGGCGCATATAGTGGCTTACAAAGGATAATGTGC 2718







NFOKEAFDRYAEMNEVVDHDTFOTCRHSOBYCTGRVYVIAALHSIYDIPADEFGAALL  
 RKNVVCYAAHPSEHSENLLEDHSHVNLDELINACFORGDGDLTFSPASSTLNYSHSYN  
 ILKVCKTYPASNREVMKEFLVTRVNTWFECKFSRDTDFLLYKGVAHKGVDSEQFYK  
 AMEDAHYKTLANCSERLLEDDSSVNVNFFPORDMVIIVPLFDIISLETSKRTRKEV  
 LVSQDFVYVNLNITVYQAKALTYSNLVSFVESIRSVINGVTARSEWDVSKLSQS  
 LSTFFFLTKLAVLDDLLISKFPALGPKTVSHQVMDISLAFGNAPFSIKERLINRKL  
 IKITENALEIRVPLVYTFHDLRSEYKMSVDMPLDIRKQWERTTEMYNALSLSVL  
 KNSDKFDVDFVSCQSLVDPMPTAAKVIIVAMNESGLTLTFEQPTEANVALAQDS  
 EKASDGLVMTSRDVEEPSIKGSMARGELQLAGSDVPSESYTRSEIEISLEQFHW  
 TASLSLHKKMSIYVTPGLKQVQWKNFIDSLVASLSAAVNLKILDKTAIDLETRQ  
 KFGVLDVASKRLVLPKAKNHAWGVVETHARKYHVALLEHDEFIITCDNWRVAVES  
 ESVVYSMAKRLTLRLIKDGEHPVSAKVVLDVPGCGKTEILSRVNFPEDLILV  
 PGROAEMIRRRANASGIIIVATKONVTRVDSFLPMNYKGARCPKRLFIDEGMLHTG  
 CVNLFVMSLCDIAYVYDQIIPYINRVTFGPYPAHFAKLEVDVEFTRTLTKCPAD  
 VTHFLNRYEGHVNCTSEKSKVSQEMVSGAASINPVSKPLGKILITFTQSDKEALIS  
 RGYADVHTVHEVQGETYADVSLVRLTPTPVSIIARDSPHVLVLSRHTKSLKYTVVM  
 DPLVSIITRDLERSVSSYLLDMYKVDAGTQXQLOVDSVFNFNFLFAAPKTDGIDSMQFY  
 YDKCLPQNSLTNNYNDWTKLTDIISLVKDCILDMGSKSVAAPKDVSLYLLKPKRKNPF  
 EMPQDTGLLENLAVIKNFNSPELSGVVDIENTASLVVDKFDVSLYLLKPKRKNPF  
 SLFRESINRWIAKQKOVITIGOLADFPVLPADVOYRHMIAQPKQKLDLSIQTEYP  
 ALQIVVTHSKKINAIQPLFSELTRQLDSDISRSRFLFTFKTKPAQIDFFGDLDSSH  
 PMVLELDVSKYDSQNEHFCABEYIWRRLGLEDFLAEVWKQGHRTTLKQYTAGIK  
 TCMYQKRGSDVTFIIGNVTIIASCLASMLPMEKLIKGAFQGDSDSLYFPKGCYBPD  
 QOANLMNPFALKPKQYGYFCGRYVTHDRGCIYVYDPLKLSKLGAKHIDKWDHL  
 EEFERSLCDVAESLNNCAIYITQLDDAVGEVHKTPPGSFVYKSLVLYSDKVLFRSLF  
 LDGSSC"

CDS

/note="unnamed protein product; 130 K protein"  
 /codon\_start=1

/protein\_id="CAA26082.1"

/db\_xref="GI:62129"

/db\_xref="GOA:P03587"

/db\_xref="Swiss-Prot:P03587"

/translation="MAYTQTATSSALLETVRGNNLTVNDLAKRLYDVAIDFNARD  
 RPKVNFSEVSEETGLATKAYPEFOITFYNTQNAVHSLAGLSLSLELYLMMQIPY  
 SLTYDIGNFASHLTKRAYVHCOMPNDVRDIRHEGOKDSIELYLSRLERGNKIPG  
 NFOKEAFDRYAEMNEVVDHDTFOTCRHSOBYCTGRVYVIAALHSIYDIPADEFGAALL  
 RKNVVCYAAHPSEHSENLLEDHSHVNLDELINACFORGDGDLTFSPASSTLNYSHSYN  
 ILKVCKTYPASNREVMKEFLVTRVNTWFECKFSRDTDFLLYKGVAHKGVDSEQFYK  
 AMEDAHYKTLANCSERLLEDDSSVNVNFFPORDMVIIVPLFDIISLETSKRTRKEV  
 LVSQDFVYVNLNITVYQAKALTYSNLVSFVESIRSVINGVTARSEWDVSKLSQS  
 LSTFFFLTKLAVLDDLLISKFPALGPKTVSHQVMDISLAFGNAPFSIKERLINRKL  
 IKITENALEIRVPLVYTFHDLRSEYKMSVDMPLDIRKQWERTTEMYNALSLSVL  
 KNSDKFDVDFVSCQSLVDPMPTAAKVIIVAMNESGLTLTFEQPTEANVALAQDS  
 EKASDGLVMTSRDVEEPSIKGSMARGELQLAGSDVPSESYTRSEIEISLEQFHW  
 TASLSLHKKMSIYVTPGLKQVQWKNFIDSLVASLSAAVNLKILDKTAIDLETRQ  
 KFGVLDVASKRLVLPKAKNHAWGVVETHARKYHVALLEHDEFIITCDNWRVAVES  
 ESVVYSMAKRLTLRLIKDGEHPVSAKVVLDVPGCGKTEILSRVNFPEDLILV  
 PGROAEMIRRRANASGIIIVATKONVTRVDSFLPMNYKGARCPKRLFIDEGMLHTG  
 CVNLFVMSLCDIAYVYDQIIPYINRVTFGPYPAHFAKLEVDVEFTRTLTKCPAD  
 VTHFLNRYEGHVNCTSEKSKVSQEMVSGAASINPVSKPLGKILITFTQSDKEALIS  
 RGYADVHTVHEVQGETYADVSLVRLTPTPVSIIARDSPHVLVLSRHTKSLKYTVVM  
 DPLVSIITRDLERSVSSYLLDMYKVDAGTQ"

CDS

/note="unnamed protein product; 30 K protein"

/codon\_start=1

/protein\_id="CAA26083.1"

/db\_xref="GI:62130"

/db\_xref="GOA:P03584"

/db\_xref="Swiss-Prot:P03584"

/translation="MALVVGKVNINFEIDLSKSEKLLPSMFTPVKSVMSKVVDKIMV  
 HENESLSEVNLKGVKLIIEGGYVCLVGLVSGEWNLPDNCRGVSVGVCKDIMERADE  
 ATLSGYTAIAKRFQPKVFNPGYITTKDAKNTQWLVIKNYKMSAGYCPLSLEFV  
 SVCIYKNNIKLGLERKVSVDGGMELSEVVDPEMNVPMVSRILAFKRTKSSKRG  
 PKNNNLGKRGSGRPKPKSPDEVEKEFDNLIEDEAETSVAADSDSY"

CDS

/note="unnamed protein product; coat protein"

/codon\_start=1

/protein\_id="CAA26084.1"

/db\_xref="GI:62131"

/db\_xref="GOA:P03576"

/db\_xref="Swiss-Prot:P03576"

/translation="MSYSITSFSPQVFLSSVWADPIELLNVCTNSLGNQFQOQARTT

VQOQFSEBWKPFPPQSTVRPFGDVYKVVYNAVLDPLITALIGAFDTRNRILIEVENQOS  
 PTTATLDAIRVDDATVARSAINNLVNLVRCGTGLYNQNTFESMGLSWTSAPAS"

## ORIGIN

Query Match 41.4%; Score 2631; DB 14; Length 6384;  
 Best Local Similarity 64.2%; Pred. No. 0;  
 Matches 4093; Conservative 0; Mismatches 2230; Indels 54; Gaps 7;  
 Qy 18 CGACACACACATTTAAACAAACAAACACATATTACAAACAAACAAACAAACAACTCTCG 77  
 Db 23 CAACACACACAAACAAACAAACATTTACATTTTACATTC-----TACAACTACAATGGCAT 78  
 Qy 78 ACATACATCTTATTAATTAGCAACGCCCTTCTTTGAAAGCGTGAGTGGTGTAAAAACACTCTCG 137  
 Db 79 ACACACAAACAGCCACATCTGTCGCCCTTTGCTTTGAGACCGTCCGAGGTHAACATACCTTGG 138  
 Qy 138 TTAATGACCTTTGCAAGAGCGCATGTATGATACGCGCGGTGGAAGAAATTTAAACGCCCGCG 197  
 Db 139 TCAACGATCTTGCACAAAGCGGCTCTATATGACACAGCGGTGATGAATTTAATGCTAGGG 198  
 Qy 198 ACCGTAGACCAAGGTCAACTTTTCCAAAACCTATTAGGAAAGAGCAACCGCTTCTAGTCT 257  
 Db 199 ACCGACGCGCTTAAAGTCAATTTTTCCAAAGTAGTAAGCGAAGAAACAGACGCTTATTGCA 258  
 Qy 258 CCAACGCGTACCCGAGGTTCACAGATTACCTTTTATAATACCTCAAAATGCCGTACACAGTT 317  
 Db 259 CCAAGCGCTACCCAGAAATTCCAAATTAATCTTCAACACAGCAATGCTGTGCAATCCC 318  
 Qy 318 TGGCTGGAGGTTTGAGAGCATTAGAATTTGGAATATCTGATCTCAAGTTCCTTATGGAT 377  
 Db 319 TTGCAGGCGTCTCCGATCATTAGAAATTTGGAATATCTGATGATGCAATTTCCCTACGAT 378  
 Qy 378 CGCCGACATATGATATAGTGGGAACTTTTGACGACACATTTGTTTCAAGGCGAGGATACG 437  
 Db 379 CATTGACATATGATATCGAGGTAAATTTTGCATCTCATCTCTTCAAGGCGGAGCATACG 438  
 Qy 438 TGCATTCCTGTATGCCCAACTCTGACATACAGATATAATGAGGACGACGACAAACAAAGG 497  
 Db 439 TTCACTGCTGTATGCCGAATCTGATGTCGCGACATAATTCGGGACGAGGCGGCAAAAGG 498  
 Qy 498 ACTCAATTGAGATGTAATTTGTCCAGATTTCTCGTCTTAAACAGGTAATTTCTCGAGTTTC 557  
 Db 499 ACAGTATTTGAATATACCTTTCTAGGCTCGAGAGGGGCAACAAACATGTCCCAAACTTC 558  
 Qy 558 AAAGGAGGCTTTTAAACAGGTATGCAAGCTCCCAACGAAGTCTGCTGCTCTTAAACTT 617  
 Db 559 AAAAGGAAGCTTTTCACAGATACGCTGAAATGCGCAACGAAGTAGTCTGTCAACGATCTT 618  
 Qy 618 TTCAAGATTGTCGAATACATCCGACAGATAGTGTGTAGAGATACGCTGCTGCTCTGTC 677  
 Db 619 TCCAAACGCTGTAGGCATTTCTCAAGAAATGTTTACACGGGAAGAGTGTATGCTATTGCTTTC 678  
 Qy 678 ACAGTTGTATGATATTTCTGTGATGATGTTGAGGCTGCGTAAATATCTAAGAAATATAC 737  
 Db 679 ATAGTATATAGATATACCTTCCGACGAGTTCCGGCGGGGCACTGCTGGAAGAAGATGTAC 738  
 Qy 738 ATGTATGTTATGACAGCTTCCATTTTGGCAGAAGCATTTATTACTAGACACAGACGAGGTTA 797  
 Db 739 ATGTATGTTATGTCGCTTTCCACTTTTCCGAGAATTTACTTCTCGAAGATTTACACAGTCA 798  
 Qy 798 CGCTTAATGAATAGGCGCACTTTCAAAGAGAGAGTGTATGATGTTCTTTTCTTTTCTTTG 857  
 Db 799 ACCTCGACGAGATCAATGCGATGTTTTCAAAGAGATGAGACAGGTTGAGCTTTTCTCTTTG 858  
 Qy 858 CTGATGAAGTACTTTTAAATTTATAGTCATAAATCAAAAATATCTTGCATTTATGTAGTTA 917  
 Db 859 CATCTGAGAGTACTCTTTAATTTATAGTCTATAGTTATTCTTAATATTCTTAAGTATGTTTCA 918  
 Qy 918 AATCTTACTTTCTCTGCTTCTAGTAGAATAGTTTACTTTTAAAGGAATTTTCTAGTCACTAGG 977  
 Db 919 AAACCTACTTCCCAGCCTCTAATAGAGAGGTTTACATGAAGAGGTTTTTTTAGTAACAGAG 978  
 Qy 978 TTAATACTTGGTTTGTAAATTTTACAAAGTAGATACCTATATTCTGTGTAAGAGTGTATA 1037

Db 979 TTAATACCTGTTTGTAAATTTCTAGAAATAGATACCTTCTTATTTGTACAAAGGTGTAG 1038  
Qy 1038 GACAAAGTAGGTGTAGATGATCAGTTCTATGAGGCGATGAAAGCCCTTTGCTTACA 1097  
Db 1039 CGCATAGGGGTGTAGATAGTGAAGAGTTTTCACAGGCTATGGAAGCCATGGCACTACA 1098  
Qy 1098 AGAAACCTTGGCCATGTTTCAACACTGAAAGAGCAATCTTTAGAGACACGGCTTCGGTTA 1157  
Db 1099 AABAGACTCTTGGCGATGTGCAACAGTGAAGAAATCTTGTAGAGGATCTTTCATCAGTTA 1158  
Qy 1158 ACTTTTGGTTCCTTAAGTAGAAGAGACATGTTAGTACCGGTGTTTGAAGGTTCTATTA 1217  
Db 1159 ATTACTGTTTCCAAAATAGGGATATGGTATGTTTCCATATTTGACATATCTCTCG 1218  
Qy 1218 CCAGCAAAAGATGACAAAGGATGAGTCAATTTGTAATCGTGACTTCCTTTACACAGTGC 1277  
Db 1219 AGACTAGTAAGAAACACGCAAGAGGTCTTAGTTTCAAGACCTTTGTTTATACAGTGT 1278  
Qy 1278 TTAATCATATCAGAACATATCAAGCCAAAGCGTTAACTTACAGAACGTAATATCTTTCG 1337  
Db 1279 TAAATCATCTCGTAGCTACCAAGGCCAAAGCGCTTACTTACTCCAAACGTATATCTTCG 1338  
Qy 1338 TGGAGTCTATAGATCCCGGTGATPAATCAATGTTGTTACTGCTAGGTCTGAATGGATG 1397  
Db 1339 TCGAATCAATTCGTTTCGAGAGTGATCATTAACGGGTTTACTGCTAGGTCTGAGTGGATG 1398  
Qy 1398 TAGATAAGCAATCTTCAACCTTGTCAATGACTTCTTCTTCGAGACTAAGCTGGCTG 1457  
Db 1399 TCGATAAATCATATTAACAGTCTTGTGATGACGTTCTTCTCAACACCAAGCTTGGC 1458  
Qy 1458 CGCTTCAAGACGATATAGTAATGCGAAAGTTTCGGTGTCTGGATAGACCACTTCTGAAC 1517  
Db 1459 TTCTGAAGACGATCTTTTGTATGACAGTTTGCATTTGGACCAAAACTGTCTCAAC 1518  
Qy 1518 TTATTTGGATGAGTGGGCAAAATTTTGTGAAACGTTTTCCTCCCACTATCAAGAGAT 1577  
Db 1519 ATGTGTGGATGAGATTTCCCTAGCTTTCGGCAATGCTTTCCTCATCGATCAAGGAAAT 1578  
Qy 1578 TGGTGACGAGGAAATCTCGATGTAGTGAAGTCTCTGAAGATCAAGATCCAGATC 1637  
Db 1579 TGATAAACCGGAAATCTGATCAAAATTCAGAGAAATGCGTTAGAGATCAGGTGCCGATC 1638  
Qy 1638 TGATGTGCATGGAAGACAGTTTCGTAGCTGAAATACACCAAGTCTGAGGAGTTACCGC 1697  
Db 1639 TTATGTCACTTCCATGATAGTTTCTGAGTACAAATGTCAAGTGGACATGCCG 1698  
Qy 1698 ATCTAGATATCAAGAGGCTTAGAAGAGCTGAGCAATGTACAGCGGTTATCAGAT 1757  
Db 1699 TGCTAGACATTAAGGAAAGATGGAAGAACTGAGGAAATGTACAAATGCACTGTCCGAAC 1758  
Qy 1758 TATCTATCCTTAAGGTGCTGATTAATTTGATATCGAAGTTCAAGACATGTGCAAGG 1817  
Db 1759 TGCTGTACTTAAATTCAGAACAGTTCGATGTTGACGCTTTTCCAGATGTGCCAT 1818  
Qy 1818 CTTTATAGTGTAGTCTGATGTGGCAGCACAGTAACTCGTTCAGTGGCCGAGATAGAA 1877  
Db 1819 CTTTATAGTCCGATCCATGACTGCGACCAAGGTAATAGTACGAGTTATGACCAACGAG 1878  
Qy 1878 GCGGTTTAACTCTTATTTGATAGCCAAACGAGGAGATGTGGCTTAAGGCTCTTAA 1936  
Db 1879 GTGGTCTTACTCTACGTTTGAACAGCCACCGCAAGCTAATGTGGCTAGCATTTGCAAG 1938  
Qy 1937 -----AGCACGGCTGAGGCGGTGTATGTTGAAACCGACATCCGAAGAGGTGAACG 1991  
Db 1939 ATCTGAAAGGCTTCTGATGGGCGTGTGGTAGTTACCTCAAGAGATGTGAGAACCGT 1998  
Qy 1992 TAAATAAAATTTCTATGCTGAGAAAGGAGATGTCCTGTGTGTGAGAAAGTCAATGTT 2051  
Db 1999 CATAAAGGTTTCGATGCGCGGTGTGAGTTACAAATGGCCGATTTATCTGGCGAGTTC 2058  
Qy 2052 TGACGAATGCTAACTTAGAGCACCGAGGTGAGTCCCTCAACGATTTCCATAGGCTT 2111

Db 2059 CTGAATCTTCATACACTAGGAGCGAGGATTTAGTCTCTCGACAGTTTTTCATATGGCAA 2118  
Qy 2112 GCGTGGATAGTGTGATTAACAAAGCAATGGCATCGGTTGTCTACACTGGCTCACTCAAG 2171  
Db 2119 CAGCTAGTTCTGTTAAATTCATAAGCAGATGTCTTCGATCGTGTACACGGGCCCTCTTAAAG 2178  
Qy 2172 TTCAACAAATGAAGAACTATGTGACAGGTTTGGCAGCTTCGTTGTCGCCCACTGTATCAA 2231  
Db 2179 TTCAACAAATGAAGAACTTTATAGACACCTGGTAGCCTCTCTCTCGGTTGTGCA 2238  
Qy 2232 ATCTATGCAAGTCACTAAAGGATGAAGTCGGGTATGATTTCTGATTTCCAGGAGAAAGTTG 2291  
Db 2239 ATCTATGCAAGTCACTAAAGGATGAAGTCGGGTATGATTTCTGATTTCCAGGAGTTG 2298  
Qy 2292 GTGTTTGGATGTCACTTTGAAAGTGGCTCTCAAACTGGCGGCCAAAGGTCAATTCAT 2351  
Db 2299 GAGTTCTGATGTTGCTTCGAAAGGTTGGCTAGTTAAACCATCCGCAAGAACCATGAT 2358  
Qy 2352 GGGGAGTTGCTCGATTTACAAAGGGGAAATGTTTACTGCACCTTCTATGTAAGAGAG 2411  
Db 2359 GGGGGTTGTTGAGACTCATCGAGGAAATATACGTCGCATTTACTGGAGCAGATGAT 2418  
Qy 2412 ATAGAATGCTGACCTGAGAGCGACTGGAGGAGGCTGTGTATCATCTGATACAATGGTAT 2471  
Db 2419 TTGCAATTTATTCGTCGATTAACCTGGCAGCGGTGGCTGTGAGTTCTGAGTCGGTAGTAT 2478  
Qy 2472 ATTCGATATTTGCAAAAGCTCCAAATCTGAGGAAACAAATGAGAGACCGTGAACCCACG 2531  
Db 2479 ATTCGATATGGCTAAACTCAGGACTCTGAGAAAGATTTGCTCAAGATGGAGAACACACG 2538  
Qy 2532 AACCTACTGCAAGATGTTACTTGTGATGGGTGGCTGTTGTGGAAGTACAAAGAG 2591  
Db 2539 TTGATTCAGCAAGGTTGTTTGGTGGATGGCTTCCAGGGTGGCGGAGACAAAGGAA 2598  
Qy 2592 ATTTTGAAGATTTGATCTGATGAGATTTGATCTGTTCTCTCGAAAAACAAGCTGCTG 2651  
Db 2599 TTCTTCGAGAGTTAATTTGAGAGAGATCTAATTTCTGTCCTGCTCAAGTCCG 2658  
Qy 2652 CTATGATCAGAGAGAGGCTAATTCATCTGATGATGATGAGGACCAATGGAATGGA 2711  
Db 2659 AGATGATCAGAGAGAGCTAATGCTGGGCATTAATAGTGGCTCAAAAGGATAATGTGC 2718  
Qy 2712 GAAAGTATGATCTTCTAATGATCCAAACCGGATCAC-----ACAAGAGGC 2762  
Db 2719 GCACGCTCGATTCATTTTGTGATTAATACGGGAAAGGCGACGCTGTCTGATTTCAAAGAT 2778  
Qy 2763 TTTTATTTGATGAAAGGTTGATGCTGCACACCGGTTGTGTTAACTTCTCTGCTGCTTATCT 2822  
Db 2779 TGCTATAGACGAGGTTTGTGCTGCATCTGCTGTTGTGTTGTTGTTGTTGTTGTTGTTGTT 2838  
Qy 2823 CTGTTTGGACATCCGATACATTTACGAGATPACACAGAGATTCCTTTTCATTAACAGAG 2882  
Db 2839 CTCTGCGATATTGCATATGTTTATGAGACACCCAAACAGATTCGTTACATCAACAGAG 2898  
Qy 2883 TTCAGATTTCCGTTATCCCAACATTTTGAAGCTGCAAGTGAAGTGAAGTGAAGTGA 2942  
Db 2899 TAACTGGTTTCCGTTACCTCGACACTTTTGAAGATTTGAGGTCGACGAGTTCGAACAA 2958  
Qy 2943 GGAGACCACTGAGATGCCAGGTGATGTAATTTTTCCTCAATCAAGTGAAGTGAAG 3002  
Db 2959 GAAAGACTACTCTCTGTTGTCGGCTGATGTACACACTTCTCTAATCAAGATGAAG 3018  
Qy 3003 GAGCGGTGACAAACCTTCAACTGTACAAAGATCGGCTCTCTCTGAGATGATAGGCGGTA 3062  
Db 3019 GACAGTAAATGTGACGCTCTCTGAAAGAAATTCAGTTTCCAGGAAATGGTTAGTGGGG 3078  
Qy 3063 AGGAGTACTAAACAGTGTTCCAACCACTAAAGGGGAAATTTGATCTTTCACCTCAGG 3122  
Db 3079 CTGCGCTTATCAATCTGTTGTCAGGCGCTTAAAGGGGAAATTTTGAATTTTTCACAGT 3138  
Qy 3123 CTGATAAATTTGAGTTAGAGGAGAGGCTTATAAGAAATGTGAAACCGTTTCATGAGATCC 3182  
Db 3139 CTGACAGGAGGCGCTTCTCTCAAGGGGCTACGAGATGTCCATCTGTACATGAGGTAC 3198





```
Db 5356 GGCCTACTGCCCTTGTCTCATTAGAAATTTGTCTGTGTCTATTGTTTATAAAAAATAATATA 5415
Qy 5400 AGAAAGGTTTGAGGGAACGTAATTTTGTAGTGTGACAGAGGCTGCCAAATTTGAACCTACT 5459
Db 5416 AAATTTGGGTTTGAGGGAAGAAATACGAGTGTGTGAACGATGGAGGACCCATCGAACTTTTCG 5475
Qy 5460 GAAAGGTTTGTAGGAGTTCGTGATGAAGTACCAATGGCTGTGGAACCTCGAAAGGTT 5519
Db 5476 GAAGAAGTTGTGTAGTTCATGAGAAATGTTTCCATGTGCGTTAGACTCGCAAGTTT 5535
Qy 5520 CCGGAAAAACAAAAAATGTTAGTAAATATGTTTAATAAT----- 5561
Db 5536 CGAACCAATCTCAGAAAGAGGTCGGAATAATAATAATTTAGGTAAAGGGCGTTCA 5595
Qy 5562 ---AAGAAAAATAATAACAGTGTGAAGAGGTTTTTAAAAATGAGGAAATTCAGAGTAAT 5618
Db 5596 GCGGGAAGGCTTAACCAAAAAAGTTTGTATGAAGTTGAAAAAGAGTTTGTATAATTTGATT 5655
Qy 5619 GTAAGTATGACGAGTCTATCGCGTCATCGAGTACGTTTAAATCAATGATGCTTATACAA 5678
Db 5656 GAAGATGAAGCGGAGACGTCGTCGCGGATTCGTATCGTATTAATAATATGCTTACTCAA 5715
Qy 5679 TCAACTCTCCGAGCCAAATTTGTTACTTATCTTCCGCTTACGCGATCTCTGCGAGCTGA 5738
Db 5716 TCACCTCTCCATCGCAATTTGTGTTTGTCTATCTGTATGGCTGACCTATAGAATTTGT 5775
Qy 5739 TCAATCTGTGTACAAATGATTTGGTAAACCGATTTCAACGCAACAGTAGGACAAACAG 5798
Db 5776 TAAAGGTTTGTAACAAATTCGTTAGTAAACAGTTTCAACACAGCAAGCAAGAACTACTG 5835
Qy 5799 TCCACACGCAATTTCCGATGCTCGGAACCTGCGCTAGTATGACAGTACGATTTCTCTG 5858
Db 5836 TTCAACAGCAGTTTACGAGGAGTGTGGAACCTTTCCCTCAGACGACCGTCAGATTTCTCTG 5895
Qy 5859 CATCGGATTTCTATGTGTATAGATATAATTTCCAGCGCTTGATCCGTTGATCAGCGGTTAT 5918
Db 5896 GCGATGTTTATAGGTGTACAGGTACATGCAATGCAATTTAGATCTCTAATTTACTGCGTTGC 5955
Qy 5919 TAAATAGCTTTGATCTAGAAATAGAAATAATAGAGTTGATTAATCAACCCGACCCGATA 5978
Db 5956 TGGCGGCTTTTGATCTAGGAATAGAAATAATCGAAGTAGAAACACGACAGAGTCCGACAA 6015
Qy 5979 CTACTCAATCTGTTAACCGACTCAGAGGTAGACGATGCTACTGATCTAGCTATAAGGGTT 6038
Db 6016 CAGCTGAACCGTTAGATGCTACCCGAGGTTAGACGCGCTACGGTTGCAATTCGGCTCTG 6075
Qy 6039 CAATCAATAATTTGGCTAATGAATGTTTGGGAACTGGCAATGTTCAATCAAGCAGGCT 6098
Db 6076 CTATAATAATTTAGTTAATGAACTAGTAAGAGGTACTGGACTGTACAAATCAGAACTTT 6135
Qy 6099 TTGAGCTGCTAGTGACTTGTCTGACCAACAACTCCGCGTACTTAGCTATTGTTGTGAG 6158
Db 6136 TTGAAAGTATGCTGCGGTTGTTCTGACCTCTGCACTCTGCACTTTAAATGCAATAGTGTCT 6195
Qy 6159 ATTTCTTAAATAAGTCCTGAAGACTTAAATTCAGGCTGCTGATACCAAAATCAGC 6218
Db 6196 GAAATATAAA-----GTTTGTGTTCTTAAACACACGTTGCTAGTACGATAACGTTAC 6247
Qy 6219 AGTGGTTGTTCTGCTCACTTAAATAATAACGATTTGTCTATCTCGATCCCAACAGTTAAACA 6278
Db 6248 AGTGTTTTTCCCTCCACTTAATCGAAGGTTAGTGTCTTGGAGCGCGGAGTAAACATA 6307
Qy 6279 TGTGATGGTATGATGTTGGTATGCGGTAAACATCGGAGAGGTTTCGAATTCCTCCCTTAA 6338
Db 6308 TATGGTTTATATATGTTCCGTAGGACGTAAGAAAAAGCGAGGATTCGAATTTCCCGGAA 6367
Qy 6339 CCGCGGTAGCGGCCA 6355
Db 6368 CCCCAGGTTGGGCCCCA 6384
```

AB083196  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

FEATURES  
source

CDS

CDS

AB083196 6385 bp RNA linear VRL 25-MAR-2004  
Tomato mosaic virus genomic RNA, complete genome,  
strain:Lila-Fukushima.  
AB083196  
AB083196.1 GI:29539243  
Tomato mosaic virus  
Tomato mosaic virus  
Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.  
1  
Yamamoto,H., Abe,T., Ueda,K., Inoue,M. and Matsumoto,T.  
A missense mutation in Tomato mosaic virus Lila-Fukushima genome  
determines its symptomless systemic infection of tomato  
J. Gen. Plant Pathol. 68, 385-389 (2002)  
2 (bases 1 to 6385)  
Abe,T., Ueda,K., Yamamoto,H., Inoue,M. and Matsumoto,T.  
Direct Submission  
Submitted (05-APR-2002) Hideki Yamamoto, Akita Agricultural  
Experiment Station, Department of Biotechnology; Higashi 1-1,  
Ogata, Akita 010-0442, Japan (E-mail:yamamoto@pref.akita.lg.jp,  
Tel:81-185-45-3101, Fax:81-185-45-3105)  
location/Qualifiers  
1. .6385  
/organism="Tomato mosaic virus"  
/mol\_type="genomic RNA"  
/strain="Lila-Fukushima"  
/db\_xref="taxon:12253"  
73. .4923  
/note="readthrough stop codon"  
/codon\_start=1  
/transl\_except=(pos:3421..3423,aa:OTHER)  
/product="180K protein"  
/protein\_id="BAC67678.1"  
/db\_xref="GI:29539244"  
translation="MAYTQTATSSALLETVGNNTLVNDLAKRRLYDTAVDFNARDR  
RPKNFVKVSEETLIATKAYPEFQITFYNTONAVHSLAGLSLEYLEMMQIPYG  
SLTYDIGNFASHLPKRAYVHCMPNLDIMRHEGQKDSIELYLSRLRGKHVP  
NFOKEADRYAEMNEVVCCHDTFOTCRHSQBCYTRGVYAIALHSYDIPADEFGAALL  
RKNVHCVAAPHSENILLESNDVLEINACFORDDRLTFSPASESTLANSYSYN  
ILKVKCTYFPASNRKVEFLVTRVNTWFKFSRDTFLLYKGVAKHGVDSEQFYK  
AMEDAHYKTLAMCNSERILLEDSVVNFWPKMRDWIIVLEFDLSIETSKRTKEV  
LVSKDFVTVLNIHTTQAKALTSNVLSFVSRVINGVATPASEWDVKSLLQS  
IKITENALEIRVPLVYTFHRLVSEYKMSVMPVLDIRKMEETEMYNALSELVNL  
LSMFFLTKLAVLKDILLISKFALGPKTVSHQVWDEISLAFGNAPPSIKERLNKL  
KNSKFDVDFVSCQSLSDPMTAAKVIIVAMNESGLTTFEOPTENVALALQDS  
EKASGALIVTSRDEEPIIKGSMARGELQAGLSDVPESYRSEIESLEQFMA  
TASLIHKQMCISIVTGPLKVOQMKNFIDSLVASLSAAVSNLKLKIDTADLDLSTRQ  
KFGVLDVASKEWLKPSAKHAGVETHARKYHVALLEHDFEFGITCDNWRRAVSS  
ESVYVSDMAKRLTLRLKDGEPHVSSAKVLDGVPGCGTKTEILSRVNEEDILV  
PGRAAEMIRRRANAGIIVATKDNVTVDSFLMNYGKRACQPKRLFIDEGLMLHTG  
CVNLFVMSLCDIAYVYGTQIIPVINEVTGPPYPAHFAKLEVDSEVETRTTLRCPAD  
VTHUNQYEGHVMTCSSEKSSQBMVSGAASINPVSKPLGKGLITFTQSDKEALLS  
RGYADVHTVHQVQGETADVSLVRLTPFVSIARDSPHVLVLSLRHRTKSLKITYTVM  
DPLVSIIRDLSRVSYLLDMKYDAGTOXQDQSVFFKNFLFVAAPKPTGDISDWQFY  
YDKCLPGNSTLNNYDVTMKLTDISLAVKQCLDMDSKSAAPKQVPTLIPMVRTAA  
EMPRGTGLENIIVAMIKENFSPVVDIENTASLVVDKFDYDLKKEKPKNFK  
SLFRESLNIAKQEQVTIQLADEDFVLDPAVDQYRMLKAQPKQLDLSIOETEP  
ALQITIVYHKKINAIQEFPLFSELTKQLDSDISSRFLFTRKTPAQIEDFFGDLDSHV  
PMDLELVSKYDQSNQEFCAVEYIWRRLGLEDFAEVMWQGHKRTTKLQDITAGIK  
TCLVYQRKSGDVTFPIGNVTIASCASLMPEKLIKAGFCQDSDLLYFPKCEYDPI  
QOANLMMNFAPKPKQGYFCGRYIVHHRGCTIVYDPLKLSKLAGAKHDKWDHL  
BEFRESLIDVAESLNNCAVYITQLDDAVGEVHTKAPPGSFVYKSLVKYLSLDFRSLF  
LDGSSC"  
73. .3423  
/codon\_start=1  
/product="130K protein"  
/protein\_id="BAC67679.1"  
/db\_xref="GI:29539245"  
translation="MAYTQTATSSALLETVGNNTLVNDLAKRRLYDTAVDFNARDR  
RPKNFVKVSEETLIATKAYPEFQITFYNTONAVHSLAGLSLEYLEMMQIPYG  
SLTYDIGNFASHLPKRAYVHCMPNLDIMRHEGQKDSIELYLSRLRGKHVP  
NFOKEADRYAEMNEVVCCHDTFOTCRHSQBCYTRGVYAIALHSYDIPADEFGAALL

```

RKNVHCYAA PHSNLLDHSVNLDEINACFORDCDRLTFSPASBSTLNSHSYNS
ILKYCKTYFPASREYVMEKEPIVTRVNTWFECKFSRIDTFLLYKGVAKHGVDSQFYK
AMEDAHYKTLACNSENILLESSVNYMFPKORDMVIIVLPFDISLETSKTRKEV
LVSDFYVYTLVNLHRTYQAKALTYSNVLSFVESRSRVIINGVTARSEWVDKSLLOS
LSMTFFLTLAVLKADLLISKALFKPTVSHQVMDDEISLAFGNAPFSIKERLINRKL
IKTENALEIRVPLXYVFDHRLVSEYKMSVDMPLDIRKQWBEETMYNALSLSVL
KNSDKFDVDFSCQSLVDPMYAAKVIIVAMNESGLTLTFEQTPEANVALALODS
EKASDGLIVTSRDEBPSIKGSMARGELQLAGSDVPESSTYTRSEBISLEQFPHA
TASLIHKQMSIIVYTGPKVQKMFIDSLVASLSAAVSNLVKLDATAIDLETRQ
KFGVLDVASKRMLVKPKAHQWGVETHARKYHVALLEHDEFGIITCDNRRVAVSS
ESVYVDMALRLLRLLKOGEPHVSAKVVLVDGPGCGKTEILSRVNEEDLIIV
PGROAAEMIRRRANAGLIIVATKDNVRTVDSFLMNYGKRARCQKRLFIIDSLMLHTG
CVNELVMSLCDIAYVGDTCQIIPYINRVTFPPYPAHPAKLEVDEVERETRLCPAD
VTFELNRYGHEVQWCTSEKKSQSEWMSGAASINPVSPLKGLITFTQSDKGLIS
RGYADVHTVREHGGTADVSLVRLTPTPVSIIARDSPHVLVLSRHTKSLUKYTYVM
DPLVSIIRDLREVSYYLLDMYKVDAGTQ"
4907..5701
/codon_start=1
/product="30K protein"
/protein_id="BAC67680.1"
/db_xref="GI:29539246"
/translation="MALVVGKVINBFIIDLSKSEKLPSMPTPVKSMVSKVDKIMV
HENESLSEVNLKGVKIEGGYVCLVGLVSGEWNLPDNCRGVSGVCMVDKMERADE
ATLGSYTAARAKRFPQKVPYNYGIITTKDAEKNIWOVLVNTKNVMSAGYCPLSLEFV
SVCIVYKNNIKLGLREKVTSVNDGPMELSEEVDEBPMENVPMSVRLAKFKTKSKRG
PKNNNLGKSGRGPXKPFDEVEKEFDNLIIDREASTVADSISY"
5704..6183
/codon_start=1
/product="17.5K protein"
/protein_id="BAC67681.1"
/db_xref="GI:29539247"
/translation="MSYSITSFQFVFLSSWADPIELLNCTNSLGNQFQOQARTT
VQOQFSEVWKFPPQSTFPGDVYKYRYNAVLPLITALIGAFDTRRIIEVENQOS
PTTAETLDAIRRVDDATVAIRSAINNVLNVELVRGLYNTQNTFESMSGLVWTSAPAS"

```

## ORIGIN

```

Query Match      41.4%; Score 2631; DB 14; Length 6385;
Best Local Similarity 64.2%; Pred. No. 0;
Matches 4093; Conservative 0; Mismatches 2230; Indels 54; Gaps 7;

Qy 18 CGACAAACAATTAACAAACAAACATATTACAAACAAACAAACAAACAAATGGCAC 77
Db 24 CAACAAACAACAACAACAACATATTACATTTTACATTC-----TACAACATACATGGCAT 79

Qy 78 ACATACATATTAATTAAGCAACGCCCTCTTGAAGCGTGAGTGGTGAACAACTCTCG 137
Db 80 ACACAAACAGCCACATCGTCGCTTGTGTTGAGACCGTCCGAGGTAAACATCTTGG 139

Qy 138 TTAATGACCTTGAAGAGGCGATGTACGATACGCGCGTGAGAAATTTAACGCCCGCG 197
Db 140 TCAACGATCTTGAAGCGCGGTCTATATGACACAGCGCGTGGATGAATTTAATGCTAGG 199

Qy 198 ACCGTAGACCAAGTCAACTTTTCCAAAACCTATTAGCGAAGACCAACGCTTCTAGTCT 257
Db 200 ACCGAGCGCTTAAGTCAATTTTCCAAAGTAGTAAGCGAAGAACAGACGCTTATTGCAA 259

Qy 258 CCAACGCGTACCGGAGTTCAGATTAACCTTTTATAATCTCAAAATGCCGTACACAGTT 317
Db 260 CCAAGCGTACCAGAAATCCAAATTAATCTTACACACAGCAATGCTGTGATTCGCC 319

Qy 318 TGGCTGGAGGTTTGAGACATTAGAATGGGAATATCTGATGCTACAAGTTCCTTATGGAT 377
Db 320 TTGACGCGGTCTCCGATCATTAGAAATGGGAATATCTGATGATGCAAAATTCCTACCGAT 379

Qy 378 CGCGACATATGATATAGTGGGACCTTTTGAGACATTTTGAGACATTTTGAGGAGGATACG 437
Db 380 CATTGACATATGATATCGAGGTAAATTTTGATCTCATCTGTTCAGAGGCGAGATACG 439

Qy 438 TGCATTGCTGTATGCCCATCTGACATACAGAGATATAATAGGCACGAGACCAAAAGG 497-
Db 440 TTCACTGCTGTATGCCGAATCTGGATGTGCGGACATATGCGGACGAGGCGCAAAAGG 499

Qy 498 ACTCAATGAGATGATTTTGTGCCAGATTTGTCTCGTTCTTAAACAAGGTAATTCCTGAGTTTC 557
```

Db 1580 TGATAAACCGGAAACTGATCAAAATTACGAGAAATCGCTTAGAGATCAGGGTGCCCGATC 1639  
Qy 1638 TGTATGTCACATGGAAGACAGAGTTCTGATGCTGAATACACCAAGTCTCAGAGATTACCGC 1697  
Db 1640 TTTATGTCACATTTCCATGATAGTTAGTTCTGAGTACAAATGTCAATGTGACATGCCCG 1699  
Qy 1698 ATCTAGATATCAAGAAAGACTTAGAAGAGCTGAGCAAAATGTACGACCGTTATCAGAA 1757  
Db 1700 TGCTAGACATTTAGGAAAAAGATGGAAGAACTGAGGAAATGTACAAATGCATGCTCCGAAC 1759  
Qy 1758 TATCTATCCTTAAGGGTCTGATTAATTTTCGATATCGCAAGTTCAAAGACATGTGCAAG 1817  
Db 1760 TGTCTGTAATTTAAAAATTCAGACAGTTTCGATGTTGACGTTTTTTCCAGATGTGCCAAT 1819  
Qy 1818 CTTTAGATGTTAGTCTGATGTGGCAGCAGAGTAATCGTTGCGAGTGCAGGCCGAGAAATAGAA 1877  
Db 1820 CTTTAGAAGTCCGATCCAATGATCTGAGCAAGAGGTAAATAGTAGCATTTATGACAAAGGAGA 1879  
Qy 1878 GCGGTTTAACTCTTACTTTTATAGCAACCGAGGAGAATGTGGCTAAGGCTCTTAAA- 1936  
Db 1880 GTGGTCTTACTCTACAGTTTGAACAGCCACCGAAGCTAAATGTTGCGCTAGCATTTGCAAG 1939  
Qy 1937 -----AGCACGGCTCTGAGCGGTGATGTTGTTGAACGAGATCCGAAAGAGTGAACG 1991  
Db 1940 ATTCTGAAAAGGCTTCTGATGGCGGTTGATAGTTACCTCAAGAGATGTTGAGGAAACCGT 1999  
Qy 1992 TAAATAAAATTTCTATTCGTGAGAAAGGAGATTCGCTGTGTGTCGAGAAAGTCAATGTT 2051  
Db 2000 CCAATNAGGTTTCGATGCGCCGTGTGAGTTTACAATTCGCCGATTAATCTGGCGACGTTTC 2059  
Qy 2052 TGACGAATGCTAACTTAGAGCACCAAGGAGTTGGAGTCCCTCAACGATTTCCATTAAGGCTT 2111  
Db 2060 CTGAATCTTCATACACTAGGAGCGAGGAGATTGAGTCTCTCGAGCAGTTTCATATGGCAA 2119  
Qy 2112 GCGTGGATGTTGATTAACAAGCAATGGCATCGGTTGTCTACACTGCTCACTCAAG 2171  
Db 2120 CAGCTAGTTCTGTTAAATTCATAGCAGATGTGTTCGATCGTGTACACGGGCCCTCTTAAAG 2179  
Qy 2172 TTCAACAAATGAAGAACTATGTGACAGTTTGGCAGCTTCGTTGTCCGCACTGATCA 2231  
Db 2180 TTCAACAAATGAAGAACTTATAGACGCTGTGAGCTCGCTCTCTGCTGGGTGTGCA 2239  
Qy 2232 ATCTATGCAAGTCACTAAAGGATGAAGTGGGTATGATTTCTGATTTCCAGGGAGAAAGTTG 2291  
Db 2240 ATCTAGTGAAGATCTTAAAGATACAGCTCGGATTTGACCTTGAAACTCGTCAAAAGTTCCG 2299  
Qy 2292 GTGTTTGGATGTCACTTTGAAAAGTGGCTCTCAAACTTCGCGCCCAAGTCAATTCAT 2351  
Db 2300 GAGTTCTGGATGTTCTCGAAAAGGTGGCTAGTTAAACCATCCCGCAAGGACCATGCA 2359  
Qy 2352 GGGGAGTTGCTCTGGATTAAGGGGAAAATGTTTACTGCACTTCTATCTTATGAAGGAG 2411  
Db 2360 GGGGGTTGTTGAGACTCATGCGAGGAATATCAGTTCGATTAATGAGGACGATGAAT 2419  
Qy 2412 ATAGAATGGTGAATGAGACGACTCGAGAGGGTGGCTGTATCATCTGATACAAATGGTAT 2471  
Db 2420 TTGGCAATTAATACGTGCGATACTGGCAGCGGTGGCTGTGAGTTCTGAGTGGTAGTAT 2479  
Qy 2472 ATTCTGATATTCGAAGTCCAAATCTGAGGAAACATGAGACGCTGAAACCCACG 2531  
Db 2480 ATTCTGATATGCTAACTCAGGACTCTGAGAAGATTTGCTCAAGATGGAGAACACACAG 2539  
Qy 2532 AACCTACTGCAAGATGTTGATGTTGGATGGGTCCTGTTGTGGAAAGTACAAAGGAG 2591  
Db 2540 TTAGTTACGAAAGTTGTTTGGTGAATGGGTTCCAGGGTTCGGGAGACAAAGGAA 2599  
Qy 2592 ATTTTGAAGATTTGATCTTGATGAGGATTTGATCTTGTTTCTTGAAAAAACAAGTGTG 2651  
Db 2600 TTCTTTTCGAGAGTTAATTCGAAGAAGATCTAATTTCTTGTCCCTGCTCAAGCTGCCG 2659  
Qy 2652 CTATGATCAGAGAGGCTAATTCATCTGAGTGTGATGAGCCCAATGGAATGCA 2711  
Db 2660 AGATGATCAGAAGAGAGCTAATTCGTCGGGCATTAATAGTGGCTACAAAGGATTAATGTC 2719

Qy 2712 GAACGGTAGATTCTCTTAATGCAATCCAAACCGCGATCAC-----ACAAGAGC 2762  
Db 2720 GCACCGTTGATTCTTTTGTGATTAATACGGGAAAAGGCGACGCTGTCTCAGTTCAAAAGAT 2779  
Qy 2763 TTTTATTTGATGAAGGGTTGATGCTGCACACCGGTTGTGTTAACTTCTCTGGTCTTACT 2822  
Db 2780 TGTTCATAGACGAAGTTTGTATGCTGCATATCTGTTGTGTGAAATTTTGTGTTGAAATGT 2839  
Qy 2823 CTGTTGGCAGCATCGCATACATTTTACGAGAGATACACAGCAGATTCCTTTCAATTAACAGAG 2882  
Db 2840 CTCTGTGGATTTGCAATATGTTTATGGAGACACCCACAGATTTCCGTACATCAACAGAG 2899  
Qy 2883 TTCAGAAATTTCCCGTATCCCAAAATTTTGAAGAGCTGCAAGTGGATGAAGTGTGATGA 2942  
Db 2900 TAACTGGTTTCCCGTACCCTGCACACTTTTGCAAAATTTGGAGGTGCAGAAAGTCGAAACAA 2959  
Qy 2943 GGAGACACACTGAGATGCCAGGTGATGAAATTTTTTCTTCAATCGAAGTACGAG 3002  
Db 2960 GAAGAACTACTCTTCTGTTGTCGGCTGATGTGCACACTTCTCTAAATCAAAAGGTATGAAG 3019  
Qy 3003 GAGCGGTGACAAACCACTTTCAACTGTACAAACGATCGGTCTCATCTGAGATGATAGCGGTA 3062  
Db 3020 GACAGTAAATGTGACGCTCTTCTGAAAAGAAATCAGTTTCCAGGAAATGTTAGTGGG 3079  
Qy 3063 AGGAGTACTAAACAGATGTTTCCAAACCACTAAAGGGGAAAATTTGAACTTCTACTCAGG 3122  
Db 3080 CTGCGTCTATCAATCTCTGTGTCCAAGCGCTTAAAGGGAAAATTTTGAATTTTCAACAGT 3139  
Qy 3123 CTGATAAATTTGAGTAGAGGAGAGGCTATAGAAATGTAACACCGTTCATGAGATCC 3182  
Db 3140 CTGCAAGAGGCGCCCTTCTCTCAAGGGCTACGCAAGATGTCCATCTGTACATAGGTAC 3199  
Qy 3183 AAGGAGAACTTTGAAGATGTGCTGCTGCTGAGATTCAGCGCACTCCACTGACTGCA 3242  
Db 3200 AAGTGAAGCTATGACAGCGTATCGTTAGTTTGCATTAACCTACGCTGTATCTATCA 3259  
Qy 3243 TTTCCAGTCTTCTCCCGCATGTTCTAGTCTGCTGACTGACTAGACACACAAAGAGCTTCAAT 3302  
Db 3260 TCGCAAGAGACAGTCCGCAATGTTCTGTCTCGTGTCAAGACACACAAAAATCCCTAAAAGT 3319  
Qy 3303 ATTACACCGTAGTGTAGATCCTTTAGTACAGAAATTTAGTATTTGCTTCTTTAAAGCT 3362  
Db 3320 ACTACCGTTGTGATGATCCTTTAGTTAGTATCATTTAGAGATTTAGAACGGTTAGTA 3379  
Qy 3363 CTTCTTTTGAAGATGATATGTTAGAGCAGGTTAGTAGCAATTTACAGATGGATG 3422  
Db 3380 GTTACTTATTAGACATGTACAAAGTAGATGAGGTACTCAATAGCAATTCAGGTCGACT 3439  
Qy 3423 CAGTGTCAAAGGTCATTAATCTTTTGTGGGAAACCTTAATTCAGGAGACTTTCAGATC 3482  
Db 3440 CTGTGTTTAAAAATTTCAATCTTTTGTAGCAGCTCCAAAGACTGGAGATATATCTGATA 3499  
Qy 3483 TACAGTTCTATTCAGATGTATGCTCCTGTTGTAATAGTACTATATCTTAACAGATGATG 3542  
Db 3500 TGCAATTTTACTATGATGATGTTCTTCTGGAACAGCAGCTGTTGTGAACAACACTACGAG 3559  
Qy 3543 CTGTTTACCATGAGGTTTACGTGATAATAGTCTTAATGTGAAGGATTTGTTCTTGAATTTT 3602  
Db 3560 CTGTTTACCATGAAATTTGACTGACATTTCTCTGAATGTCAAGATTTGCATATTAGATATG 3619  
Qy 3603 CCAAAAGTATTCGATGCGAAAGGAGGTGAAACCAATGTCTAGAGCCAGTTTTCGTTACCG 3662  
Db 3620 CTAAGTCTGTAGTCTCTCGAAAGATGTCAAACCACTTTAATACCGATGGTACGAAACCG 3679  
Qy 3663 CGGCGGAACCCCAAGGCTGCGAGCTACTCGGAAATCTGCTGCAATGTTTAAAGAA 3722  
Db 3680 CGGCAAGAAATGCTCCGACAGCTGAGCTGTGGAATCTAGTGGATGATTAAGAA 3739  
Qy 3723 ATTTCAAACGACACAGACCTGACGGGAGGATTTGCAATTTGAGAGCAGCCGATCTGTTGTAG 3782  
Db 3740 ATTTTAAATTCACAGAGTTGTCCGGAGTAGTTGATATTGAAATATCTGCTATCTTTAGTGG 3799

QY 3783 TAGATAAGTTTTTGTAGTCTATTTTATTAAGAAAGAAATATACACAAAAATATTGCTG 3842  
Db TAGATAAGTTTTTGTAGTCTATTTTACTTAAGAAAGAAAGAAACCAACAAAAATTTT 3859  
QY 3843 GAGTGATGACGAAGATTCAATGATGAGATGGTGGAAAAACAGGAAAGAGTACTATTGG 3902  
Db CACTGTTTAGTAGAGAGTCTCTCAATAGGTGGATAGCAAGCAAGAACAGTCACAATTG 3919  
QY 3903 ACGACTTGGCTAACTACAAATTTTACAGATCTGCGGCGCATCGATCAGTACAGACACATGA 3962  
Db GTCAAGTTGGCCGATTTTGAATTTTGTGATCTTCAGCCGTTGATCAGTACAGGCATATGA 3979  
QY 3963 TCAAGGCTCAACCAAAAAACAGAAATTTGACCTTTTCAATTTCAAGAAAGTAACTCCTCTGC 4022  
Db TTAAGCGCAACCGAGCAGAAATCTGGATCTGTCAATTCAGACAGAAATATCCAGCGTTGC 4039  
QY 4023 AAACAATTTGTCTACCATTTCAAGCAGATCAACGGTATTTTGG---COGGTTTCTCAGAGC 4079  
Db AAACGATTGTGTATCAATTCAGAAAGAAATCAACGCAATATTGGTCTCTTTTCAGTGAGC 4099  
QY 4080 TTACAGGTTGCTGCTCGAGGCAATTTGAATCTAAGAAAGTTTCTTTCTTTACTAGGAAAA 4139  
Db TTACAGGCAATTAATTTGACAGTATTTGACTCAAGCAGATTTCTTCTTTACGAGAAAGA 4159  
QY 4140 CTCCAGAACAGATTCAGAAATTTTCTCGATCTCGACTCGCAGTTTCTTCTATGATGTT 4199  
Db CACCGCTCAGATCGAAGATTTCTTCGGAGATCTAGACAGTCATGTCCCAATGACATAC 4219  
QY 4200 TAGAACTGGATATTTCTAAGTATGATAAGTCAAGAACAGAGTTTCATTTGTCTGATAGT 4259  
Db TTGAGTTGGATTTTCGAGTATGATTAAGTCTCAAAACGAGTTTCATTTGTCTGTTGAT 4279  
QY 4260 ATGAAATATGGAAGATTTGGTCTCAATGATGTTTTTGGCGAAGTGTGGAACAAGGCG 4319  
Db ACGAAATCTCGAGGAGACTGGGTCTCGAGGATTTCTTGGCAGAGTGTGGAACAAGGCG 4339  
QY 4320 ACAGGAAACAACTTTGAAGATTTACATTTCTGCGAATCAAGACATGCTGTGGTATCAA 4379  
Db ATAGGAAACCACTCTGAAGATTTACATCTGCTGTGTAATAAAACGTGTTATGTGTACAGA 4399  
QY 4380 GGAAGCGGTGATGCTACTTTTCATCGCAATCTGTTATATAGCAGTCTGCTTGG 4439  
Db GAAAGATGTTGATTTTACAACTTTTATCGGTATACCGTCATCTGCTTCTGTTCTAG 4459  
QY 4440 GTTCAATGTTTACCGATGGAAGAGTCTATAAGGTGCTTTTGTGGAGACGATTTCCGTTT 4499  
Db CATCAATGCTCCCGATGGAAGAAATTTGATAAAGGAGCCTTCTCGGAGAGTACAGATTTGT 4519  
QY 4500 TGTATTTTCCAAAGGTTTGGATTTCCCTGACATTCAGTCAATGCTCAATCTCATGTGA 4559  
Db TGTACTTTTCCTAAGGGTTGTGAGTATCCCGATATACAAACAGCTGCTAATCTAATGTGGA 4579  
QY 4560 ATTTTGAGGCCAACTGTATAGAAGAGGTACGGTACTTTTGTGGTAGATACATCATAC 4619  
Db ATTTTGAGGCCAACTGTATAGAAGAAATATGGGTACTTCTCGCGGAGGTACGTGATTC 4639  
QY 4620 ACCATGATAAGGGAGCAATAGTGTATTAATGATCTTTGAAGTTGATCTCCAAACTTGGGG 4679  
Db ATCAGATAGAGGTTGCATAGTACTACTACGACCCTTTGAAGCTGATTTGGAACCTTGGTG 4699  
QY 4680 CAAAACATATCAAGATATGATCACTTAGAAGAGTTAAGGTTCTTTTGTGCGATGTTG 4739  
Db CTAACACATCAAGGATTTGGATCAATTTGGAGGAGTTTCAAGATCCCTCTGTGATGTTG 4759  
QY 4740 CTTGTTCCGCTCGGAACTGGTCTTAGGCTTTCCGAGCTGAACGCACTATCAAGAGG 4799  
Db CTGAGTCTGTT---GAACAATTTGGCGGTATACACAAATTTGACGACGCTGTTGGGAGG 4816  
QY 4800 TTCAATAAAACCGCGATTTGATGGTTGGTTGCTTTTAAATTTGTTTAAACAAATTTTGTGTG 4859  
Db TTCAATAAAACCGCCACCTGGTTGGTTGTTTATTAAGAGTTTAAAGTATTTGTGTCAG 4876  
QY 4860 ATAAATTTTATTTAGAACTTTGTTTAAATGCTGTAGTCTCAGAGATCTGTCTCAAA 4919

Db 4877 ATAAAGTTTTTGTAGAGTTTATTTCTTGTGCTCTAGTTGTTAAAGGTTAAGTAAAT 4936  
QY 4920 ATTACGAGTTCAATGATCTTTTGGAAACAGAGATGAGATCTTCCGGCATTCATGACTAAG 4979  
Db ATTAATGAGTTTATCGATCTGTCAAGTCTGAGAAACTTCTCCCGTGCATGTTCAAGCCT 4996  
QY 4980 GTCAAGAGTCTTAAAGATATCGATCTGCGAAGAGATGAGTCTGTTAAGATGATAGTCTT 5039  
Db GTAAAGAGTGTATGGTTTCAAGGTTTGAAGATATGTTGTCATGAAAAATGAATCATG 5056  
QY 5040 TCTGATCTGATATTACTTAAAGGTTTAAAGTTAGTTAGTAAAGAGGTTATGTTGCTTAGCT 5099  
Db TCTGAGTAAATCTCTTTAAAGGTTGTAAGAACTTATAGAGGTGGTATGTTGTTGTTAGTC 5116  
QY 5100 GATTGTTAGTGTCTGGGAGTGGAACTCTCCGGATAACTGCCGTGTGGTGTGTCAGTGT 5159  
Db GGTCTTGTGTCGGTGTGAGTGGAAATTTACCAGATAATTTGCCGTGTGGTGTGAGTGTG 5176  
QY 5160 TGTATTTAGATAGAGAAATGAAAAGAGTAAAGAAACAGCTGGGTGGGTATCAGCC 5219  
Db TGCATGTTTGAAGAGAAATGGAAGAGCGGACGAAACACACTGGGGTCTATATTACACT 5236  
QY 5220 CCTGCTTGCAAAAGAAATTTTCTTTTAAAGCTAATCCTTAATTTCAATAACATCCGAG 5279  
Db GCTGCTCTAAAAGCCGTTTTCAGTTTAAAGTGTGCCAACTACGGTATTATTACAAAG 5296  
QY 5280 GATGCTGAGAAAGCACCGCTGGCAAGTGTATTAGTGAATATCAAGAGGAGTGGCTATGGAAGAA 5339  
Db GATGCAGAAAGAACATATGCGAGTCTTAGTAAATATTAATAATGTAATAAATGAGTGC 5356  
QY 5340 GGATATGCTCTTTATCTTTGGAGTTCGTTTCAATTTGTGTAGTACATAAATAATATGTA 5399  
Db GGTCTACTGCTTTTGTCAATTTAGAAATTTGTCTGTGTATTTGTTTATAAAAAATATA 5416  
QY 5400 AGAAAGGTTTGGGGAACGTTATTTTGGAGTGTGACAGACGCTCGCCAAATTTGAATCTACT 5459  
Db AATTGGGTTTGGAGGAGAAAGTAAACGAGTGTGAACGATGGAGACCCATGGAATTTCA 5476  
QY 5460 GAAAAGGTTTGTGAGGAGTTTCGTGGATGAAAGTACCAATGGCTGTGAAACTCGAAAAAGTT 5519  
Db GAAGAAGTCTGTGATGAGTTTCATGAGAAATGTTCCAAATGCGGTAGACTCGCAAGTTT 5536  
QY 5520 CCGGAAAAAAGAAAGTGGTAGGTAATAATGTTTAAAT----- 5561  
Db CGAACCAATCTCTCAAAAGAGGTTCCGAAAAATAATAATAATTTAGGTAAAGGGGCTTCA 5596  
QY 5562 ---AAGAAATAATAACAGTGTGTAAGAGGTTTAAATAATGAGGAAATTTGAGGATAAT 5618  
Db GCGGAAAGGCTTAAACCAAAAGTTTGTGAAGTTGAAAAGAGTTTGAATTTGATTT 5656  
QY 5619 GTAAGTATGACGAGTCTATCGGCTCATCGAGTACGTTTAAATCAATATGCTTTATACAA 5678  
Db GAAGATGACGCGAGAGCTCGTCCGCGATTCGTATTAATAATGTTCTTACTCAA 5716  
QY 5679 TCAACTCTCCGAGCAATTTGTTTATCTTCCGTTTACGAGATCTCTGTGAGCTGA 5738  
Db TCACTTCTCCATCGCAATTTGTTGTTTGTCTCATCTGTATGGGCTGACCCCTATAGAATTGT 5776  
QY 5739 TCAATCTGTGTACAAATGCAATTTGGTAAACAGTTTCAACGCAACAGCTAGGACACAG 5798  
Db TAAACGTTTGTACAAATTCGTATAGGTAAACAGTTTCAACGCAACAGCAAGAACTACTG 5836  
QY 5799 TCCAAACAGCAATTTGCGGATGCTCGGAAACCTGTGCTAGTATGACAGTGAAGATTTCTGT 5858  
Db TTCAACAGAGATTTCAGCGAGGTGTGGAACCTTTCCCTCAGAGCACCGTCAAGATTTCTGT 5896  
QY 5859 CATCGGATTTCTATGTTGATATATAATTCGACGCTTGTATCCGTTGATCAGCGGTTAT 5918  
Db GCGATGTTTATAAGTGTACAGGTACAAATGTCAGTTTGTAGATCTCTTAATTTACTGCTGTC 5956  
QY 5919 TAAATAGCTTTGATCTAGAAATAGAAATATAGAGTTTGAATCAATCAACCCCGACCGAATA 5978

Db	5957	TGGGGCTTTCGATATAGTAATAAGTAATCGAAGTAGAAAAACGACAGTCCGACAA	6016
Qy	5979	CTACTGAATCGTTAAACGAGCTCAGAGGTAGACGATGCTACTAGTACTATAGGCTT	6038
Db	6017	CAGCTGAACGTTAGATGCTACCCGACGGGTAGACGACGCTACCGGTTGCAATTCGGTCTG	6076
Qy	6039	CAATCAATAATTTGGCTTAATGAATCGTTCCTGGAACTGGCATCTTCAATCAAGCAGGCT	6098
Db	6077	CTATAATATATTAGTTAATGAATAGTAGAGGTACTGGACTGTACAATCAGATACTT	6136
Qy	6099	TTGAGACTGCTAGTGGACTTGTCTGGACCAAACTCCGGGCTACTAGCTATTGTTGTGAG	6158
Db	6137	TTGAAAGTATGCTCGGGTTGGTCTGGACCTCTGCACCTCGCATCTTAAATGCATAGGTGCT	6196
Qy	6159	ATTTCCTAAATAAAGTCGCTGAAGACTTAAATTCAGGGTGGCTGATACCAAAATCAGC	6218
Db	6197	GAATATATAA-----GTTTGTGTTCTTAAAAACACACGCTGGTACGATACGATAC	6248
Qy	6219	AGTGGTTGTTCCGCTCACTTAATAATACGATTGTTCATATCTGGATCCAAACAGTTAAACCA	6278
Db	6249	AGTGGTTTCCCTCACTTAATCGAAGGTAGTGTCTTGGAGCGCGGAGTAAACATA	6308
Qy	6279	TGTCATGGTGTATCTGTGTTATCGGCTGAAACATCGGAGAGGTTTGAATCCTCCCTTAA	6338
Db	6309	TATGGTTTCATATATGTCGTAGGACGCTGAAAAAAGCGAGGATTTCGAATTCCTCCCGGAA	6368
Qy	6339	CCGCGGTAGCGGCCCA	6355
Db	6369	CCCCCGGTGGGGCCCA	6385

RESULT 8	TM0132845	6384 bp	RNA	linear	VRL 24-FEB-1999
LOCUS	tomato mosaic virus genes encoding RNA-dependent RNA polymerase, movement protein and coat protein.				
DEFINITION					
ACCESSION	AJ132845				
VERSION	AJ132845.1 GI:4456886				
KEYWORDS	coat protein; movement protein; RNA-dependent RNA polymerase.				
SOURCE	Tomato mosaic virus				
ORGANISM	Tomato mosaic virus				
REFERENCE	1 Zhou, X., Xue, C., Chen, Q., Qi, Y. and Li, D.				
AUTHORS	Complete nucleotide sequence of a Chinese isolate of tomato mosaic virus				
TITLE	Unpublished				
JOURNAL	2 (bases 1 to 6384)				
REFERENCE	Zhou, X.				
AUTHORS	Direct Submission				
TITLE	Submitted (06-FEB-1999) Zhou X., Zhejiang University, Huajiachi, Institute of Biotechnology 268 Kaixuan Road, Hangzhou, Zhejiang 310029, CHINA				
JOURNAL	Location/Qualifiers				
FEATURES	1. .6384				
source	/organism="Tomato mosaic virus"				
	/proviral				
	/mol_type="genomic RNA"				
	/isolate="S-1"				
	/db_xref="taxon:12253"				
	72. .4922				
	/note="180kDa protein				
	amber termination codon				
	/codon_start=1				
	/transl_except=(pos:3420..3422,aa:OTHER)				
	/product="RNA-dependent RNA polymerase"				
	/protein_id="CAB36997.1"				
	/db_xref="GI:4456887"				
	/db_xref="GOA:O9YKD6"				
	/db_xref="Swiss-Prot:O9YKD6"				
	/translation="MAYTQTATSSALLETIRGNNTLVNDLAKRRLYDTAVDFNARDR				
	RPKNFNSKVSSEQTATKAYPEFQITFYNTQNAVHSLAGLSLELYLMMQIPYGR				
	SLTYDIGNFASHLFKGRAYVHCCMPNLDVRDIMRHGQKDSIELYLSRLERGNKHVP				

CDS

```

NFOKEAFDRYAEMPNVCHDTFQTCGHSQECYTRGVYATIALHSIYDI PADEFGAALL
RKNVYCYAFHSESNLLSHVNDLEINACFORQDGRDLTFSPASSTLNYSHYSN
ILKYVCKYTPASNREVMKEFLVTRVNTVFQKFSR IDTFLLYKGVAHKGVDSEQFYK
AMEDAHYKKTILAMCNSERILLEDSSSNVWFPPKMDWVI VPLFDISETSKRKEV
LVLDKDFYVTLNHI RTYQAKALTYSNVLSVESIRSRVINGVTARSDWDKSLLOS
LSMTFFLHI KLAVLKDILLISKFALGPKTVSQHWDEISLAFNAFSPISKERLNRLK
IKITENALEIRVPLDYVTFHDLVSEYKMSVMPVLDIRKMEETEEMYNALSELVL
KNSDFDVDFVSCQSLVEDPMTAAKVI VAMNSESGLTTFEQPTANVALALQDS
EKASDGLAVTSDRVEBPSIKSGMARGELQAGLSGDVPESSTRSEISLEQPHMA
TASLIJHKMCSI VYTGPKVQOMKNFIDSLVASLSAAVSNLVKILKDTAAIDLETRQ
KFGVLNVASKRWLVPSAKNHAGVETHAKYHVALLEHDEFGIITCDNWRRAVSP
ESVYSDMAKLTURLRLKDGEPHVSAAKVVLDGVPCGCKTEILSRVNEEDLILV
PGRQAEMIRRRANASGI VQTKDNRTVDSFLMNYGKARQCQFRLFIDEGMLMHTG
CVNPLFMSLSDIAYVYGTQOIPIYINRTGFPYPAHFAKLEVEDETRTTRCPAD
VTHFNORYEGHVMTSEKKSVSQEMVSGAASINPVSKPLKGLITLFTQSDKEALLS
RGYTDVHTVHEVOGETVADVSLVRLTPTPVSIIARDSPHVLVLSRLTKSLKYTVVM
DPLVSIIRDLERVSSYLLDMYKVDAGTQ"
4906..5700
/note="30kDa protein"
/codon_start=1
/product="movement protein"
/protein_id="CAB36997.1"
/db_xref="GI:4456889"
/db_xref="GOA:O9YKQ9"
/db_xref="Swiss-Prot:O9YKQ9"
/translation="MALVYKGVNINEFIDLSKSEKLPSPMFTPVKSVWVSKVDKIMV
HENESLSVNLKGLVKVLEGGYVCLVGVKNSGEWNLPCNCRGVSCLVDRKEMRADE
ATLGSYVTAARKRFQFQVFPVNYGITTAKDKNIWQVLVNTKNVQMSAGYCPLSLEFV
SVCIVYKNTKLGIREKVTSDNGGEMSELSVEVDEWENVPMSVRLAKFRTKSKRGK
PKNNNLGKSGRGGPKPKSFDEVEKEFDNLIEDAETSVDSDSY"
5703..6182
/codon_start=1
/product="coat protein"
/protein_id="CAB37000.1"
/db_xref="GI:4456890"
/db_xref="GOA:O41345"

```

CDS

CDS

CDS





QY 2052 TGACGAATGCTAACTTAGAGCACCAGGAGTTGGAGTCCCTCAACGATTTCCATAAGGCTT 2111  
Db 2059 CTGAATCTTTCATACACTAGGAGCAGGAGATTGAGTCTCTCGACAGCTTTTCATATGGCAA 2118  
QY 2112 GCGTGGATAGTGTGATTACAAAGCAATGGCATCGGTGTCTACACTGGCTCACTCAAAG 2171  
Db 2119 CAGCTAGTTCGTTAAATTCATAGCAGATGTGTTTCGATCGGTGATACCGGCGCTCTTAAAG 2178  
QY 2172 TTCAACAAATGAAGAACTATGTGGACAGTTTGGCAGCTTCGTTGTCGCCACATGTATCAA 2231  
Db 2179 TTCAACAAATGAAGAACTTTATAGACAGCCTGGTAGCTCTCTCTGCTGGGTGTCGA 2238  
QY 2232 ATCTATCGAAGTCACTAAAGATGAAGTCGGGTATGATCTGTGATTCAGGAGGAAAGTTG 2291  
Db 2239 ATCTAGTGAAGATCCTTAAAGATACAGCTCGATCGACCTTGAACCTCGTCAAAAGTTG 2298  
QY 2292 GTGTTTGGGATGTCACATTTGAAAAGTGGCTCTCAAACTCGGCCCAAGGTCATTCAT 2351  
Db 2299 GAGTTCCTGGATGTGCTTCGAAAAGTGGCTAGTTAAACCAATCCGGAAGAACCATGCAT 2358  
QY 2352 GGGGAGTTGTCCTGGATTACAAGGGGAAAATGTTTACTGCACTTCTATCTTATGAAGGAG 2411  
Db 2359 GGGGGGTTGTGAGACTCATCGGAGGAAATATCACGTCGCATTAATCGAGCACGATGAAT 2418  
QY 2412 ATAGAATGGTCACTGAGCGACTGGAGGAGGTGGCTGTATCATCTGATACAAATGGTAT 2471  
Db 2419 TTGCCATTTATACGTGCGATAACTGGCAGCGGTGGCTGTGAGTCTCGGTAGTAT 2478  
QY 2472 ATTCGTATATTGCAAAAGCTCCAAATCTGAGGAAACAAATGAGAGCGGTGMAACCCACG 2531  
Db 2479 ATTCGTATATTGCTTAAGTCTGAGACTCTGAGAGATTGCTCAAGATGGAGAACCAACG 2538  
QY 2532 AACCTACTGCAAGAGATGGTACTTGTGATGGGTGCTGTTGTGGAAAGTACAAAGGAG 2591  
Db 2539 TTAGTTCAGCAAGGTGGTGTGATGGATGGCTTCAGGAGTGGGAAAGAAACGAAGGAAA 2598  
QY 2592 ATTTTGAAGATTGATCTTGATGAGGATTTGATCTTGGTTCCTGGAAAACAAGCTGCTG 2651  
Db 2599 TTCTTTTCGAGAGTTAAATTTGGAAGAAGATCTAAATCTTGTGCTCCCTGGTTCGTCAGGCTGCG 2658  
QY 2652 CTATGATCAGAAGAGGGCTAAATTCATCTGACCTGATAAGAGCCACAAATGCAATGTGA 2711  
Db 2659 AGATGATCAGAGAGAGCTAAATGCGTCCGGCATATAGTGGCTACAAAGGATTAATGTGC 2718  
QY 2712 GAACGGTAGATTCACTTCTAATGATCCAAACCGCGATCAC-----ACAAGAGGC 2762  
Db 2719 GCACCGTCGATTCATTTTGTATGAATTAACGGAAGGGGCAACGCTGTCAAGTTCAAAAGAT 2778  
QY 2763 TTTTATTTGATGAGGGTGTGATGCTGCACACCGGTGTGTGTTACTTCTGCTGCTTATCT 2822  
Db 2779 TGTTCATAGACGAAGGTTTGATGCTGCATACCTGTTGTGTGAAATTTCTTGGTTGAAATGT 2838  
QY 2823 CTGTTTCGACATCGCATACATTTACCGGAGATACACAGCAGATTCCTTCAATTAACAGAG 2882  
Db 2839 CTCTGAGGATATTTGCATATGTTTATGGAGACCCCAACAAATTCGCTACATCAACAGAG 2898  
QY 2883 TTCAGAAATTTCCGTATCCCAACATTTGAGAGCTGCAAGTGGATGAAGTTGAGATGA 2942  
Db 2899 TAACTGTTTCCCGTACCTGCACTTTGCAAAATTTGGAGGTGCGAGGAAGTCGAAACAA 2958  
QY 2943 GGAGGACCACTGAGATGCCAGGTGATGTAATTTTCTTACATCGAATCGAATGACGAAG 3002  
Db 2959 GAAGAACTACTCTTCTGTTGCTCGGCTGATGTCACACACTTCTTCAATCAAGGTTATGAAG 3018  
QY 3003 GAGCGGTGACAAACCACTTCAACTGTACAAACGATCGGTCTCATCTGAGATGATAGGCGGTA 3062  
Db 3019 GACACGTAATGTGCACCTCTTCTGAAAAGAAATCGGTTTCCAGGAATGTTAGTGGGG 3078  
QY 3063 AGGAGTACTAAACAGTGTTCAAAACCACTAAAGGGGAAAATTTGTAACCTTCACTCAGG 3122  
Db 3079 CTGCGTCTATCAATCTCTGTGTCGAAGCGGTTAAAGGGGAAAATTTTGAACCTTCAACAGT 3138  
QY 3123 CTGATAAATTTGAGTTAGAGGAGGAGGCTATAGAATGTGAACACCGCTTTCATGAGATCC 3182

Db 3139 CAGACAAGGAGGCCCTTCTCTCAAGGGCTATACAGATGCTCCATCTGTATCATGAGGTAC 3198  
QY 3183 AAGGAGAGAAACCTTTGAAGATGTGCTGCTGAGTTGAGCGCAACTCCACTGACTCTGA 3242  
Db 3199 AAGGTGAGACTTATGACAGCTATCGTAGTTCGACTAACCTACGCTGTATCATCA 3258  
QY 3243 TTTCCAAAGTCTTCCCGCATGTTCTAGTCGCTCTGACTAGACACAAAGAGCTTCAAAAT 3302  
Db 3259 TCGCAAGAGACAGCTCCGCATGTTCTGCTCTCGTTGTCAAGACACACAAAAATCCCTAAAGT 3318  
QY 3303 ATTACACCGTAGTGTAGATCCCTTTAGTACAGATAATTAAGTGTCTCTTCTTAAAGCT 3362  
Db 3319 ACTACACCGTTGTAGTGGATCCTTTTAGTTAGTATCATTTAGAGATTTAGAACGGGTTAGTA 3378  
QY 3363 CTTCTCTTTTGAAGATGTATATGTTAGAACGAGTGTAGTAGCAATTTACAGATGGATG 3422  
Db 3379 GTTACTTTTATGACATGTACAAAGTAGATGACAGTACTCANTAGCAATTTACAGGTCGACT 3438  
QY 3423 CAGTGTTCAAAGTCAATAATCTCTTTTGTGGCAACACTAAATCAGGAGACTTTCCAGATC 3482  
Db 3439 CTGTGTTTAAATAATTTCAATCTTTTGTAGCAACTCCAAAGACTGGAGATATATCTGATA 3498  
QY 3483 TACAGTTCATTTACGATGTATGCTCCCTGGTAATAGTACTATCTTAAACAGTATGATG 3542  
Db 3499 TGCATTTTACTATGATAAGTGTCTTCTCTGGGAACAGCAGCTGTTGTAACCACTACGACG 3558  
QY 3543 CTGTACCATGAGGTTACGTGATAATAGTCTTAAATGGAAGGATTTGTTCTTGTGATTTTT 3602  
Db 3559 CTGTACCATGAATTTGACTGACATTTCTGATGTCAAGATTTGCAATATTAGATATGT 3618  
QY 3603 CCAAAAGTATTCGATGCCAAAGAGGATGAAACATGTCTAGAGCCAGTTTTCGGTACCG 3662  
Db 3619 CTAAGTCTGTAGCTGTCTCGAAAGATGCCAAACCAACTTTAATACCGATGGTACGAACGG 3678  
QY 3663 CGCGGNAACCGCAAGGCTGCAGGACTACTCCAAATCTGTTGCAATCATTAATAAGAA 3722  
Db 3679 CGCGAAGATGCTCCGACAGCTGGACTATTGGAATACTTAGTTGGATGATTAAGAA 3738  
QY 3723 ATTTCAACGACACAGCCTGACCGGAGCATTTGACATTTGAGAGCAGCGCATCTGTTGTAG 3782  
Db 3739 ATTTTAAATTCACCAGAGTTGTCCGGAGTAGTTGATTTGAAATACTGTCATCTTTAGTGG 3798  
QY 3783 TAGATAAGTTTTTTGATAGCTATTTTATTAAGAAAGAAATTAACAAAAAATTTGCTG 3842  
Db 3799 TAGATAAGTTTTTTGATAGTTATTTTAAAGGAAAAAGAAAAACCAAAATTTTTT 3858  
QY 3843 GAGTGATGACGAGGATTCATGATGAGATGGTTGGAACACAGAAAGAGTACTATTGG 3902  
Db 3859 CACTGTTTAGTAGAGAGTCTCTCAATAGGTGGATAGCAAGCAAGAACAGTCACAATTTG 3918  
QY 3903 ACAGCTTGGCTAACTACAATTTTACAGATCTCGCGCCATCGATCAGTACAAGCACATGA 3962  
Db 3919 GTCAGTTGGCCGATTTTGAATTTTGTGATCTTCCAGCGCTTGATCAGTACAGGCATATGA 3978  
QY 3963 TCAAGGCTCAACCAAAACGAAATTTGACCTTTCAATTCAGAAATGAATACCTGCTGCTGC 4022  
Db 3979 TTAAGCGCAACCGAAAGCAGAACTGGATCTGTCAATTCAGACAGAAATATCCAGCGTTGC 4038  
QY 4023 AACAATTTGCTACCAATTCGAACAGATCAACCGTATTTTGG--CCGGTTTCTCAGAGC 4079  
Db 4039 AAACGATTTGTATCATTTCAAGAAATCAACGCAATTTTGGTCTCTTTTCAGTGAGC 4098  
QY 4080 TTAACAAGTTTGTCTCGAGGCAATTTGATTTCTAAGAAAGTTTCTTTTCTTTTACTAGGAAA 4139  
Db 4099 TTACAGGCAATTTACTTTGACAGATTTGACTCAAGCAGATTTCTTGTCTTTTACGAGAA 4158  
QY 4140 CTCAGAAACAGATTCAGAAATTTTCTCGGATCTCGACTCGCAGCTTCTTATGATGATGT 4199  
Db 4159 CACCGCTCAGATCGAAGATTTTCTCGGAGATTTGACAGTCTGATGCTCCCAATGGACGTAC 4218  
QY 4200 TAGAACCTGGATATTTCTTAAGTATGATAAGTCAAGAACGAGTTTCAATTTGCTGTGATG 4259



Db 4219 TTGAGTTGGATGTTTCGAAGTATGATAAGTCTCAAAACGAGTTTCATTGTGCTGTTGAGT 4278  
Qy 4260 ATGAATAATGGAAGAAAGTTGGTCTCAATGAGTTTTCGCGCAAGTGTGGAACACAGGCG 4319  
Db 4279 ACGAAATCTGGAGGAGACTGGTCTGGAGGATTTCTTTGGCAGAGTGTGGAACACAGGCG 4338  
Qy 4320 ACAGGAAACCACTTTGAGGATTTACATTCCTGCAATCAAGACATGTCGTGTGATCAAA 4379  
Db 4339 ATAGGAAACCACTCTGAAAGATTAACCTGCTGTTATTAAGAGTGTGTTATGTTACAGA 4398  
Qy 4380 GGAAGAGCGGTGATGTAATCTTTTCATCGGCAATCTGTTTAATAAGACAGTCTGCTGG 4439  
Db 4399 GAAAGAGTGTGATGTTTACAACTTTTATCGTAAATACCGTCACTATTCGTTCTGCTGCTG 4458  
Qy 4440 GTTCAATGTTACCGATGGAAGAGTCTATAAGGTCGTTTTCGTTGAGACATTCGTTT 4499  
Db 4459 CATCAATGCTCCGATGGAAGAAATTTGATTAAGAGGCTTCTGCGGAGATGACAGTTTGT 4518  
Qy 4500 TGTATTTTCCAAAGGTTTCGATTTCCCTGACATTCAGTCAATGCTCAATCTCATGTGA 4559  
Db 4519 TGTACTTTCTTAAGGTTGTGAGTATCCCGATATACAAAGCTGCTAAATCTAATGTGA 4578  
Qy 4560 ATTTTGAGGCAAACTGTATAGAAAGAGGTACGTTTCTTTTGTGAGATACATCATAC 4619  
Db 4579 ATTTTGAGGCAAACTGTTTAAGAGCAATATGGTACTTCTGCGGAGGTACGTGATTC 4638  
Qy 4620 ACCATGATAAGGAGCAATAGTGTATATGATCTTTTGAAGTGTATCTCAAACTTGGGG 4679  
Db 4639 ATCAGATAGAGGTGTCATAGTATACTACGACCTTTTGAAGCTGATTTTCGAAACTTGGTG 4698  
Qy 4680 CAAAACATATCAAGGATTAATGATCATTAGAGGTTTAAGGTTCTTTTGTGCGATGTTG 4739  
Db 4699 CTAAACACATCAAGGATTTGGATCATTTGGAGGATTCAGAAATCCCTCTGTAATGTTG 4758  
Qy 4740 CTGTTGCTCGGAAACTGCTGCTTAGGCTTTCCGAGCTGAAACGCACTATCAAGGAGG 4799  
Db 4759 CTGAGTGTI -- GAACAAATGCGGTATTTACACAAATTTGGACGCTGTTGGGAGG 4815  
Qy 4800 TTCAATAAACCGGATGATGTTGTTGTTGTTTAAATGTTGTTAAACAAATTTTGTGTTG 4859  
Db 4816 TTCAATAAACCGGCTGTTGTTGTTGTTTAAAGATTTAGTAAAGTATTTGTTGTCAG 4875  
Qy 4860 ATAAATTTTATTTAGAACTTTGTTTAAATGCTGTAGTCTCAGAGATCTGTCMAA 4919  
Db 4876 ATAAAGTTTGTGTAGAAATTTATTTCTGATGCTCTAGTGTGTTAAAGGTAAGGTAAT 4935  
Qy 4920 ATTAGCGAGTTCATTTGATCTTTTGAACAGGATGAGATCTTCCGCAATTCATGACTAAG 4979  
Db 4936 ATTAATGAGTTTATCGATCTGTCAGAGTCTGAGAACTTCTCCGTCGATGTTCAAGCCT 4995  
Qy 4980 GTCAAGAGTGTAGAAATATCGACTGTGGAACAAGATTAGCTGTGTTAAGAAATGATGCTTT 5039  
Db 4996 GTAAAGAGTGTATGTTTCAAGGTTGATAAGATTATGTCCTCAATGAATAATGAATCATG 5055  
Qy 5040 TCTGATGTAGATTTACTTAAGGTTTAAGTGTAGTTAAGAAAGGTTATGTCCTTAGCT 5099  
Db 5056 TCTGAAGTAAATCTTTTAAAGGTTTAAACTTATAGAAGGTGGGTATGTTTGCCTAGTT 5115  
Qy 5100 GATTTGTGATGTCGCGGAGTGAATCTCCGCAATCTCCGTCGTTGTTGTTGTTGTTGTT 5159  
Db 5116 GGTCTGTTGTTGTCGCGTGTGAGTGAATTTACAGATTAATTTGCGTGTGTTGTTGAGTTC 5175  
Qy 5160 TGTATTTGTAGATAAGAGATGAAGAGGATGAAGAGCAACGCTGGGTGCTATCAAGCC 5219  
Db 5176 TGCTTGTGTTGACAGAAATGGAAGAGCGGACGAGCCACTCTGGGTCTATATACACT 5235  
Qy 5220 CTGCTGTCAAAAGAAATTTTCTTTTAAAGCTAATCCCTAATTTATCAATAACATCCGAG 5279  
Db 5236 GCTGCTGTAAAAAGCGGTTTCAGTTTAAAGTGTGTCCTCCAAATTAACGTTATTAACAACAAAG 5295  
Qy 5280 GATGCTGAGAGACACCGTGCAGGTGTAGTGAATATCAAGAGGAGTGGCTATGGAAGAA 5339  
Db 5296 GATGCAAGAAAGAACATATGCGAGGCTTTAGTAAATATTAATAATGTAATAATGATGTCG 5355

Qy 5340 GGATCTGCTCTTTATCTTTGGAGTTCGTTTCAATTTGTGTAGTACATATAAATAATGTA 5399  
Db 5356 GGCTACTGCGCTTTGTTCATTTAGAAATTTGTGTCTGTGTATTTGTTATATAAATAATATACA 5415  
Qy 5400 AGAAAAAGGTTTGAGGGAACGTAATTTTGAGTGTGACAGCGCTCGCCAAATTTGAACCTCACT 5459  
Db 5416 AATTTGGGTTTGAGGGAAGAAAGTAAACGAGTGTGAAACGATGAGGAGCCATCGAACTTCA 5475  
Qy 5460 GAAAAAGGTTTGAGGAGTTCGTTGGATGAAGTAAACCAATGCTGTGAAACTCGAAAAAGTT 5519  
Db 5476 GAAGAAAGTGTGTTGATGAGTTTCATGAGAAATGTTTCCAATGCTCGGTTAGACTCGCAAGTTT 5535  
Qy 5520 CCGAAAAACAAAAAGAAATGTTAGTAAATGTTTAAATTTAGGTAAGGGGCGTTCA 5561  
Db 5536 CGAACCAATCTCAAAAAAGAGTCCGAAAAATAATAATAATTTAGGTAAGGGGCGTTCA 5595  
Qy 5562 --- AAGAAAAATAATAACAGTGTGAAGAGGTTTTAAAAATTCAGGAAAAATGAGGATAAT 5618  
Db 5596 GCGGGAAGGCTTAACCAAAAAAGTTTGTGAAAGTTGAAAAAGAGTTTGATAATTTGATTT 5655  
Qy 5619 GTAAAGTGAACGAGTCTATCGGTCATCGAGTACGTTTTTAAATCAATATGCTTTATACAA 5678  
Db 5656 GAAAGTGAAGCCGAGAGCTCGGTCGCGATTTCTGATTCGTTTAAATAATGCTTTACTCAA 5715  
Qy 5679 TCAACTCTCGAGCCAAATTTGTTTACTTATCTTCCGTTACGAGATCTCTGTGCGAGCTGA 5738  
Db 5716 TCACTTTCTCCATCGCAATTTGTTGTTTGTCTGTATGGGCTGACCTTATAGAATTTGT 5775  
Qy 5739 TCAATCTGTGTACAAATGCAATTCGGTAAACCAAGTTTCAACGCAACAGCTAGGACAAACAG 5798  
Db 5776 TAAACGTTTGTACAAATTCGTTAGGTAAACAGTTTCAACAGCAAGCAAGCAACTACTG 5835  
Qy 5799 TCAACAGCAATTTGCGGATGCTGGAACCTGCTGCTAGTATGACAGTGAAGATTTCTG 5858  
Db 5836 TTCAACAGCAGTTTCAGCGAGTGTGGAACCTTTCCCTCAGAGCACGCTCAGATTTCTG 5895  
Qy 5859 CATCGAATTTCTATGTATAGATATAATTCGAGGCTTGATCCGTTGATCACGGCGTTAT 5918  
Db 5896 GCGATGTTTATAAGGTTGACAGTACAATGTCAGTTTGTAGATCTCTTAATTTACTGCGTTGC 5955  
Qy 5919 TAAATAGCTTTGATATCTAGAAATAGAAATATAGAGTTGATTAATCAACCCGCAACCAATA 5978  
Db 5956 TGGGCTTTTCGATCTAGGAATAGAAATATCGAGTGAAGAACCCAGCAGAGTCCGACAA 6015  
Qy 5979 CTACTGAAATTCGTAAACGCACTCAGAGGTAGACGATGCTACTGTAGCTTATAAGGCTT 6038  
Db 6016 CAGCTGAAACGTTTAGATGCTACCCGAGGTTAGACGAGCTACGGTTGCAATTCGGTCTG 6075  
Qy 6039 CAATCAATTTTGGCTTAATGAACCTGGTTGCTGGAACCTGTCATGTTCAATCAAGCAGGCT 6098  
Db 6076 CTATAAATAATTTAGTTAATGAACCTAGTAAGAGGTTACTGGAATCTGCAATCAGAAATACTT 6135  
Qy 6099 TTGAGACTGCTAGTGGACTTGTCTGGACCAACCTCCGGCTACTTAGCTATTTGTTGTAG 6158  
Db 6136 TTGAAAGTATGCTGCGGTTGCTGGACCTTCGACCTGCACTGCACTTTAAATGCTAGTGCC 6195  
Qy 6159 ATTTCCATAAATAAAGTTCGCTGAAGCTTTAAATAATTCAGGGTGTGCTGATACCAAAATCAGC 6218  
Db 6196 GAAATATAA ----- GTTTGTTTCTTAAACACACGCTGTTACGATACGATAACGTTAC 6247  
Qy 6219 AGTGGTTGTTGCTGCTCACTTAATAATATAACGATTTGTCATATCTGGATCCAAACAGTTAAACCA 6278  
Db 6248 AGTGTGTTTCCCTCCACTTAAATCGAAGGTTAGTGTCTTGGAGCGCGGAGTAAACATA 6307  
Qy 6279 TGTGATGTTGATATCTGTTGTTATGCGGTAAAAACATCGAGAGGTTTGAATCTCCCTCAA 6338  
Db 6308 TATGGTTTCTATATGTCGTTAGGACATGTAATAAAGAGGAGGATTCGATTTCCCTCCCGAA 6367  
Qy 6339 CCGCCGTTAGCGGCCCA 6355  
Db 6368 CCGCCGTTGGGGCCCA 6384

RESULT 9  
TMO243571  
LOCUS  
DEFINITION  
Tobacco mosaic virus (K1 strain) complete RNA genome.  
ACCESSION  
AJ243571  
VERSION  
AJ243571.1 GI:6572455  
KEYWORDS  
126K protein; 17.7K protein; 183K protein; 29K protein; coat protein gene; movement protein; ORF1; ORF2; ORF3; RNA-dependent RNA polymerase.

SOURCE  
ORGANISM  
Tobacco mosaic virus  
Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.

REFERENCE  
1  
Belenovich, E.V., Novikov, V.K. and Zavriev, S.K.  
Biological properties and genome structure of the Kazakh strain K1 of tobacco mosaic virus  
Unpublished  
2 (bases 1 to 6383)  
Zavriev, S.K.  
Direct Submission  
Submitted (19-JUL-1999) Zavriev S.K., Laboratory of Molecular Virology, Institute of Agricultural Biotechnology, Timiryazevskaya st. 42, Moscow 127550, RUSSIA  
Related Sequences: AF062519, Z92909.

COMMENT  
FEATURES  
source  
1. .6383  
/organism="Tobacco mosaic virus"  
/region  
/variation  
/mol\_type="genomic RNA"  
/strain="Kazakh strain"  
/isolate="K1"  
/db\_xref="taxon:12242"  
/country="kazakhstan"  
join(72..3418,3422..4922)  
/function="RNA replication"  
/note="ribosomal slippage ORF1"  
/codon\_start=1  
/product="rna-dependent RNA polymerase, 183K protein"  
/protein\_id="CAB62911.1"  
/db\_xref="GI:6572456"  
/db\_xref="GOA:Q9Q1T8"  
/db\_xref="Swiss-Prot:Q9Q1T8"  
/translation="MAYTQTATSSALLETVRGNTLVNDLAKRRLYDTAVDEFNARD  
RPKNFSKVSEEQTLIATKAYPEFOITFYNTQNAVHSLAGLSLELEYLMQIPIYG  
SLTYDIGNFASHLPFKGRAYVHCMPNLDVRDIRHEGQKDSIELYLSRLERGNKVP  
NFOKEADRYAEMNEVVCCHDTFOTCRHSQECYGRVYAIALHSIYDIPADEFGAALL  
RKNHVCYAAPHFSENILLESNDVNDINACFORDDRLTFSPASESTLVNHSYSN  
ILKVTQYTFPASNREYVTKFELVTRVNTWFCESRIDTELLYKGVAKHGVNDQFYK  
AMEDAHYKKTILAMCNSRIILLESSSVNYFFPKMRDVIPLDFISLETSKRTRKEV  
LVSKDFVYTVLNIHTYQAKALTVSNLSPVSRVIRSVINGVTARSEDWDVDSLLQS  
LSMTFFLHTKLAVLDKLLISKALPGKTVSHQVWDEISLAFGNAPSIIKERLINRKL  
IKITENALEIRVPDLYTFHDLRVSEYKMSVPMVLDIRKMEETEEMYNALSELVL  
KTSKDPVDVFSQCSLEVDPMTAAKVIVAMVNESGLTLTFEQTEANVALALQDS  
EKASDGLAVTSRDEEPSIRGSMARGELQAGLSDVPESSTYRSEIESLEQPHMA  
TASSLIHKQMSIYVTGPKVQOMKVFIDSLVASLSAAVSNLVKILKDTAIDLETRQ  
FGVLVDVASKRWLVKPSAKNHAWGVETHARKYVHALLEHDEFGLITCDNRRVAVSS  
ESVYVDMAKLRTLRLLKQGEHPVSAKVLVVDVPGCGKTKTEILSRVNFEDLILV  
PGROAEMIRRRANASGIIIVATKDNVTVDSFLMNYGKGARCOFKPLFIDEGMLHTG  
CVNFLNVEMLCDIAVYGDYQOIQIPYINRVTPGPPYPAHFAKLEVDETRRTLRCPAD  
VTHFLNRYEGHVCMCTSEKKSQSVQEMVSGAASINPVSKPLGKILITFTQSDKEALLS  
RGYADVHTVHEVQSETYADVSLVRLTPTPVSIIARDSPHVLVSLSRHTKSLKYITVM  
DPLVSIIRDLERVSYSLLDMYKVDAGTQ"  
4906..5700  
/gene="ORF3"  
4906..5700  
/gene="ORF3"  
/function="TMV movement in plants"  
/codon\_start=1  
/product="29K movement protein"  
/protein\_id="CAB62913.1"  
/db\_xref="GI:6572458"  
/db\_xref="GOA:P03584"  
/db\_xref="Swiss-Prot:P03584"  
/translation="NALVVKGVNINEFIDLSKSEKLLPSMFTPVKSVVMVSKVDKIMV  
HENESISEVNLKGVKLEGGVYCLVGLVSGEWNLPDNCRGVSVCMVDKMERADE  
ATLSYTYAAKAKRFQFKVVPNYGITTDAEKNIWQVLNLIKVMKMSAGYCLPLSEFV  
SVLSYITAAKIKLGRKETSVDNGGPMELSEKIVNDEFMENVPMVSLAKFRYKSKRG  
PKNNNLGKRGSGRGPFKPSFDEVEKEFDNLIEDAEATSVADSDSY"  
5703..6182  
/gene="coat protein"  
5703..6182  
/gene="coat protein"  
/function="viral RNA coating"  
/codon\_start=1  
/product="17.7K protein"  
/protein\_id="CAB62914.1"  
/db\_xref="GI:6572459"  
/db\_xref="GOA:Q9Q1T6"  
/db\_xref="Swiss-Prot:Q9Q1T6"  
/translation="MSYPIPSQSFVFLSSVWADPIELNLVNTSLGNQFOFQOARTT  
VQOQSEVMEFPQSTVRFPQGVYKYRYNAVLDPLITALLGTFITRNRIIEVENQRKS  
PTTATETLDAITRVDDATVAIRSAIINNLVNLVRLVGTGLYNQNTFESMSGLVWTSAPAS"

CDS  
72..3422  
/function="protein containing methyltransferase and  
RNA-helicase domains"  
/note="ORF2"  
/codon\_start=1  
/product="126K protein"  
/protein\_id="CAB62912.1"  
/db\_xref="GI:6572457"  
/db\_xref="GOA:Q9Q1T8"  
/db\_xref="Swiss-Prot:Q9Q1T8"  
/translation="MAYTQTATSSALLETVRGNTLVNDLAKRRLYDTAVDEFNARD  
RPKNFSKVSEEQTLIATKAYPEFOITFYNTQNAVHSLAGLSLELEYLMQIPIYG  
SLTYDIGNFASHLPFKGRAYVHCMPNLDVRDIRHEGQKDSIELYLSRLERGNKVP  
NFOKEADRYAEMNEVVCCHDTFOTCRHSQECYGRVYAIALHSIYDIPADEFGAALL  
RKNHVCYAAPHFSENILLESNDVNDINACFORDDRLTFSPASESTLVNHSYSN  
ILKVTQYTFPASNREYVTKFELVTRVNTWFCESRIDTELLYKGVAKHGVNDQFYK  
AMEDAHYKKTILAMCNSRIILLESSSVNYFFPKMRDVIPLDFISLETSKRTRKEV  
LVSKDFVYTVLNIHTYQAKALTVSNLSPVSRVIRSVINGVTARSEDWDVDSLLQS  
LSMTFFLHTKLAVLDKLLISKALPGKTVSHQVWDEISLAFGNAPSIIKERLINRKL  
IKITENALEIRVPDLYTFHDLRVSEYKMSVPMVLDIRKMEETEEMYNALSELVL  
KTSKDPVDVFSQCSLEVDPMTAAKVIVAMVNESGLTLTFEQTEANVALALQDS  
EKASDGLAVTSRDEEPSIRGSMARGELQAGLSDVPESSTYRSEIESLEQPHMA  
TASSLIHKQMSIYVTGPKVQOMKVFIDSLVASLSAAVSNLVKILKDTAIDLETRQ  
FGVLVDVASKRWLVKPSAKNHAWGVETHARKYVHALLEHDEFGLITCDNRRVAVSS  
ESVYVDMAKLRTLRLLKQGEHPVSAKVLVVDVPGCGKTKTEILSRVNFEDLILV  
PGROAEMIRRRANASGIIIVATKDNVTVDSFLMNYGKGARCOFKPLFIDEGMLHTG  
CVNFLNVEMLCDIAVYGDYQOIQIPYINRVTPGPPYPAHFAKLEVDETRRTLRCPAD  
VTHFLNRYEGHVCMCTSEKKSQSVQEMVSGAASINPVSKPLGKILITFTQSDKEALLS  
RGYADVHTVHEVQSETYADVSLVRLTPTPVSIIARDSPHVLVSLSRHTKSLKYITVM  
DPLVSIIRDLERVSYSLLDMYKVDAGTQ"  
4906..5700  
/gene="ORF3"  
4906..5700  
/gene="ORF3"  
/function="TMV movement in plants"  
/codon\_start=1  
/product="29K movement protein"  
/protein\_id="CAB62913.1"  
/db\_xref="GI:6572458"  
/db\_xref="GOA:P03584"  
/db\_xref="Swiss-Prot:P03584"  
/translation="NALVVKGVNINEFIDLSKSEKLLPSMFTPVKSVVMVSKVDKIMV  
HENESISEVNLKGVKLEGGVYCLVGLVSGEWNLPDNCRGVSVCMVDKMERADE  
ATLSYTYAAKAKRFQFKVVPNYGITTDAEKNIWQVLNLIKVMKMSAGYCLPLSEFV  
SVLSYITAAKIKLGRKETSVDNGGPMELSEKIVNDEFMENVPMVSLAKFRYKSKRG  
PKNNNLGKRGSGRGPFKPSFDEVEKEFDNLIEDAEATSVADSDSY"  
5703..6182  
/gene="coat protein"  
5703..6182  
/gene="coat protein"  
/function="viral RNA coating"  
/codon\_start=1  
/product="17.7K protein"  
/protein\_id="CAB62914.1"  
/db\_xref="GI:6572459"  
/db\_xref="GOA:Q9Q1T6"  
/db\_xref="Swiss-Prot:Q9Q1T6"  
/translation="MSYPIPSQSFVFLSSVWADPIELNLVNTSLGNQFOFQOARTT  
VQOQSEVMEFPQSTVRFPQGVYKYRYNAVLDPLITALLGTFITRNRIIEVENQRKS  
PTTATETLDAITRVDDATVAIRSAIINNLVNLVRLVGTGLYNQNTFESMSGLVWTSAPAS"

ORIGIN  
Query Match 41.3%; Score 2623.8; DB 14; Length 6383;  
Best Local Similarity 64.2%; Pred. No. 0;  
Matches 4095; Conservative 0; Mismatches 2227; Indels 55; Gaps 8;  
Qy 18 CGACAAACAAATTTAAAAACAAAAACAACTATTATCAAAACAAAAACAAACAAATGGCAC 77  
Db 23 CAACAACAACAACAACAACAATTTACATTTTACATTC----TACAACTACAAATGGCAT 78  
Qy 78 ACATACAACTTATAATTAGCAACGCCCTTCTTGAAGCGTGAGTGTGTAAACAACTCTCG 137  
Db 79 ACACAAACACAGACATCGTCGCTTTGCTTGACCCGTCGAGGTAAACAATACCTTGG 138  
DGSSC"

138 TTAAATGACCTTGCAGAGCGGCATGTACGATACGCGCGTGGAGAAATTTAAACGCCGCG 197 QY  
139 TCAACGACTTGTCAAAAGCGCGCTATATGACACAGCGGTGCGATGAATTTAAATGCTAGGG 198 Db  
198 ACCGTAGACCAAGAGTCAACTTTTCCAAAATATTAGCGAAGAGCAACGCTTCTAGTCT 257 QY  
199 ACCGCGAGGCTTAAGTCATTTTCCAAAGTAGTAGAGGAGAACAGACGCTTATTGCAA 258 Db  
258 CCAACGCGTACCGGAGTCCAGATTACCTTTTATTAATCTCAAAATGCCGTACACAGTT 317 QY  
259 CCAAGGCTACCCAGAAATCCAAATTTACATTTCAACACGCGAGAACGCTGTGCAATCCC 318 Db  
318 TGCGTGGAGTTGAGAGCAATTAGAAATGGAATATCTGATGCTACAAGTTCCCTATGGAT 377 QY  
319 TTGCGGCGGTCTCCGATCATATTAGAAATGGAATATCTGATGATGCAAAATCCCTACGGAT 378 Db  
378 CGCGGACATATGATATAGGTGGGAACTTTGCGAGCACATTTGTTTCAAAGGCGAGGATTAACG 437 QY  
379 CATTTGACATATGATATCGGAGGTAAATTTTGCATCTCATCTGTTCAAAGGCGAGCATACG 438 Db  
438 TGCAATGCTGTATGCCCAATCTCGACATACGAGATATAATGAGGCGAGAAAGCAAAAGG 497 QY  
439 TTCACTGCTGTATGCCGAATCTGGATGTCGCGACATATATGCGGCGACGAGGCGCAAAAGG 498 Db  
498 ACTCAATTTGAGATGATTTTGTCCAGATTGCTGCTGTTCTAAACAGGTAAATTCCTGAGTTTC 557 QY  
499 ACAGTATTGAACATATACCTTTCTAGGCTCGAGAGGGGCAACAAACATGTCCCAAACTTCC 558 Db  
558 AAAGGAGGCTTTTAAACAGGTATGCGAGAGCTCCCAACGAAAGTCTGCTGCTCTAAAACTT 617 QY  
559 AAAGGAGGCTTTGCGACAGATACGCTGGAATGCGAAACGAGAGTAGTCTGTACAGTACTT 618 Db  
618 TTCAGGATGTCGAATACATCCGCGAGAGATAGTGTGAGAGATAGCTGTTGCTCTGC 677 QY  
619 TCCAAACGCTGAGGCAATCTCAAGAAATGTTTACCGGGAAGGTATGCTATTGCTTTGC 678 Db  
678 ACAGTTTGTATGATATTTCTGTGCTGATGAGTTTGGAGCTGGTTAAATATCTAAGAAATATAC 737 QY  
679 ATAGTATATACGATATACCTTCCGCGAGGTTCGCGCGGCACTGCTGAGAAAGAAATGTAC 738 Db  
738 ATGTATGTTATGAGCTTCCATTTTGGCAGAAGCATTTATTTACTAGACACGAGGTTA 797 QY  
739 ATGTATGTTATGCGCTTTCCACTTTCCGAGAAATTTACTTCTCGAAGATTTACAGCTCA 798 Db  
798 CGCTTAATGAATAGGCGCACTTTCAAAGAGAGGTGATGATGTTTCTTTTCTTTTG 857 QY  
799 ACCTCGATGAGATCAATGCAATGTTTCCAAAGAGATGGAGACAGGTTGACTTTTTTCTTTG 858 Db  
858 CTGATGAAGTACTTTAAATATATAGTCAATAAATACAAATAATCTTGCATATGTAGTTA 917 QY  
859 CATCTGAGAGTACTTTAAATATATAGTCAATGTTATTTCTAAATTTCTTAAGTATGTTGCA 918 Db  
918 AATCTTACTTTCTGCTCTAGTAGAATAGTTTACTTTTAAAGGAATTTTGTACTACTAGGG 977 QY  
919 AAATCTTCTCCGAGCTCTAAATAGAGAGGTTTACATGAAGGAGTTTTTTAGTAACATAGAG 978 Db  
978 TTAATACTTGGTTTGTAAATTTTACCAAGTAGATACCTATATTTCTGTACAAGAGTGTTA 1037 QY  
979 TAAATACCTGTTTGTAAATTTTCTAGAATAGTACTTTCTTATTGTACAAGGTTAG 1038 Db  
1038 GACAAGTAGGTTGATGATGATGATCTATGAGGCGATGGAAGAGCCCTTTGCTTTACA 1097 QY  
1039 CGCATAGGGGTGAGACAATGAGCAGTTTTTCAAGGCTATGGAAGAGCGCATGGCACTACA 1098 Db  
1098 AGAAACCTTGGCCATGTTCAACACTGAAGAGCAATCTTTAGAGACAGCGCTTCGCTTA 1157 QY  
1099 AAAAGACTCTTGGCATGTGCAACAGTGAAGAAATCTTTGTTAGAGGATTTCTCATCAGTTA 1158 Db  
1158 ACTTTTGGTTTCCCTAAGATGAAGGACATGCTGTATGATGACCGCTGTTTGGGGTTCTATTA 1217 QY  
1159 ATTACTGTTTCCAAAATGAGGGATGATGATAGTTTCCACTATTTGACATATCTCTCG 1218 Db

1218 CCAGCAAAAGATGCAAGGAGTGAGGTCAATGCTTAATCGTGACTTCGTTTACACAGTGC 1277 QY  
1219 AGACTAGTAAAGAACACGAAAGAGGTCTTAGTTTCAAAGGACTTTGTTTATACAGTGT 1278 Db  
1278 TTAATCATATCAGAAATATCAAGCCAAAGCGTTAACTTACCAAGAAAGTATATTCTTTG 1337 QY  
1279 TAAATCACATTCGTACGTACCAGGCCAAAGCGCTTACTTACTCCAAGCGTGTATCTTTG 1338 Db  
1338 TGGAGTCTATAGATCCCGGTGATTAATCAATGGTGTACTGTAGGTCTGAATGGGATG 1397 QY  
1339 TCGAATCAATTCGTTCCGAGAGTGATCAITTAACCGGGTGTACTGTAGGTCTGAGTGGGATG 1398 Db  
1398 TAGATAAGCAATTCCTTCAACCTCTCAATGACTTTCTTCTTGAGACTTAAGCTGCTG 1457 QY  
1399 TCGATAAATCATTAATTACAGTCTCTGTGATGACGTTCTTCTTACATACAGCTTGCCG 1458 Db  
1458 CGCTTCAAGACGATATAGTAATGGGAAAGTTTCGGTCTTGGATAAAGACCACTTCTGAAC 1517 QY  
1459 TTCTGAAGACGATCTTTTGATTTAGCAAGTTTGCACCTTGGTCCAAAACCTGTCTCAAC 1518 Db  
1518 TTATTTGGGATGAGGTGGGCAAAATTTTGTGAAAACGTTTTTCCCACTATCAAGAGAGAT 1577 QY  
1519 ATGTGTGGGATGAGATTTCCCTAGCTTTCCGCAATGCTTTCCCATCGATCAAGGAAAGAT 1578 Db  
1578 TGGTGAGCAGGAAAATTTCTGGATGTAAGTGAGAAATGCTCTGAAAGATCAAGATCCAGATC 1637 QY  
1579 TGATAAACCGAACTGTATCAAAATTTACGAGAAATGCGTTAGAGATCAGGGTGCCGATC 1638 Db  
1638 TGTATGTCATGTAAGAGACAGGTTCTGTAGCTGAATACACAAAGTCTGAGGAGTTACCGC 1697 QY  
1639 TTTATGCTACTTTCCATGATAGGTTAGTTTCTGAGTACAAAATGTCAAGTGACATGCCG 1698 Db  
1698 ATCTAGATATCAAGAGGACTTAGAAGAGCTTAGCAAAATGTACGAGCGCTTATCAGAAAT 1757 QY  
1699 TGCTAGACATTTAGGAAAAGATGGAAGAACTGAGGAAATGTACAATGTCACTGTCTCCGAA 1758 Db  
1758 TATCTATCTTAAAGGTGCTGATAATTTTCGATATCCGAAAGTTCAAAGACATGTGCAAG 1817 QY  
1759 TGTCTGTACTTAAACCTCAGACAGTTGATGTTGATGTTTTTCCAGATGTGCCAAT 1818 Db  
1818 CTTTATAGTGTAGTCTGTGCGCAGCAGAGTAATCGTTGCGAGTGGCGCAGAAATAGAA 1877 QY  
1819 CTTTATAGTTCGATCCATGACTGCGAGCAAGGTATAGTAGCAGTTATGAGCAACGAGA 1878 Db  
1878 GCGTTTAACTCTTATTTGATAAGCCAAACGAGGAGAAATGTGCGTAAAGGCTTTAAA - 1936 QY  
1879 GCGGTCTTACTCTCAGCTTGAACAGCCCAAGCTAATGTTGCGCTAGCATTTGCAAG 1938 Db  
1937 -----AGCACGGCTCTGAGCGCTGATGTTCTTGAACCGACATCCGAGAGGTGAACG 1991 QY  
1939 ATTCTGAAAAGGCTTCTGACGGGCGGTGGTAGTTTACCTCAAGAGATGTTTGAAGAACCGT 1998 Db  
1992 TAAATAAATTTTCTATTGCTGAGAAAAGGAGATTTGCTGTGTGTGAGAAAAGTCAATGTT 2051 QY  
1999 CCATAAGGGTTGATGCGCGGTGGTGTAGTTACAAATTGGCCGATTTATCTGGCGACGTT 2058 Db  
2052 TGAAGAAATGCTTAACTTTAGAGCACGAGAGTTGAGTCCCTCAACGATTTCCATAAGGCTT 2111 QY  
2059 CTGAATCTTCAATACACTAGGAGCGAGAGATTGAGTCTCTCGAGCAGTTTTCATATGCA 2118 Db  
2112 GCGTGTAGTGTGATTAACAAAGCAATTCGCATTCGGTGTCTACACTGGCTCACTCAAG 2171 QY  
2119 CAGCTAGTTGCTTAAATTTCAATAGCAGATGTTTTCGATCGGTGATACAGGGCCCTCTTAAG 2178 Db  
2172 TTCAACAAATTAAGAACTATGTGACAGATTTTGGCAGCTTCTGTTGTCCGCACTGTATCAA 2231 QY  
2179 TTCAACAAATTAAGAACTTTATAGACAGCTGTGTAGCTCTCTCTGCTGCGGTGTGCGA 2238 Db  
2232 ATCTATGCAAGTCACTAAAGGATGAAGTGGGTATGATTTCTGATTTCCAGGAGGAAAGTTG 2291 QY  
2239 ATCTAGTGAAGATCTTAAAGATACAGCTGCGATGACCTTGAACCTCGTCAAAAGTTGCG 2298 Db  
2292 GTGTTTGGGATGTCACTTTTGAAAAAGTGGGCTCTTCAAACTCGCGGCCAAAGGTCATTCAT 2351 QY

Db 2299 GAGTCTCGATGTTGCTTCGAAAAGGTGGCTAGTTAAACCAATCCGCAAGAACCATGCAT 2358  
Qy 2352 GGGAGTGTCTCTGATTAACAGGGGAAAAGTTTACTGCACTTCTATCTTATGAAGAG 2411  
Db 2359 GGGGGTGTGTGAGACTCATCGAGGAAATATCATCGTCGCTATTACGAGCACGATGAAT 2418  
Qy 2412 ATAGAATCGTGACTGAGAGCGACTGGAGGGGTGGCTGTATCATCTGTATACAAATGGTAT 2471  
Db 2419 TTGGCAATTAATACGTGCGCAACCTGGCGACGGGTGGCTGTGAGTCTGAGTCTGGTAGTAT 2478  
Qy 2472 ATTCTGATATTGCAAAAGCTCCAAATCTGAGGAAAACAATGAGAGACGGTGAACCCACG 2531  
Db 2479 ATTCTGATATGCTAAACTCAGGACTCTGAGGAATGCTCAAAGATGGGAAACCAACG 2538  
Qy 2532 AACCTACTGCAAGATGTTACTTCTGGATGGGTGGCTGTGTTGGAAAGTACAAAGGAG 2591  
Db 2539 TTAGTTACGCAAGAGTGGTGTGTGGATGGCTTCAGGGTGGCGGAAAACAAGGAAA 2598  
Qy 2592 ATTTTGAAGATTTGATCTTGATGAGGATTTGATCTTGGTTCTCGGAAAACAAGCTGCTG 2651  
Db 2599 TTCTTTCGAGAGTTAAATTTGGAAGAGATCTAATCTTGTCTCCCGGTGCTCAAGCTGCCG 2658  
Qy 2652 CTATGATCAGAAGAGGGCTAATTAATCTGGAATGATGAAGAGCCACAATGACAAATGTA 2711  
Db 2659 AGATGATCAGAAGAGAGCTAATCGTCGGGCATAATAGTGGCTTACAAAGGATAATGTGC 2718  
Qy 2712 GAACGGTAGATTCACTTCTATGATCATCCAAACCGCGATCAC-----ACAAAGGC 2762  
Db 2719 GCACCGTCGATTCATTTTGTGATGAATTAACGGGAAGGGGACGCTGTCAAGTTCAAAAGAT 2778  
Qy 2763 TTTTATTTGATGAAGGGTTGATGCTGCACACCGGTGTGTAACTTCTCGTGGCTTATCT 2822  
Db 2779 TGTTCATAGACGAAGTTTGATGCTGCATACTGGTGTGTGAATTTCTTGGTTGAATGT 2838  
Qy 2823 CTGTTGGACATCGCATACATTTACGGAGATACACAGCAGATCTCTTTTCAATTAACAGAG 2882  
Db 2839 CTCTGTGCGATTTGTCATATGTTTATGGAGACACCCCAACAAATTCCTGTACATCAACAGAG 2898  
Qy 2883 TTCAGAAATTTCCCGTATCCCAACATTTTGAGAGAGCTGCAAGTGGATGAAGTTGAGATGA 2942  
Db 2899 TAACGTGTTTCCCGTACCTGTCACACTTTGCAAAATTTGGAGGTGCGAGAAAGTGCAGAACAA 2958  
Qy 2943 GGAGGACACACTGAGATGCCAGGTGATGTGAATTTTTTCTTCAATTCGAAGTACGAAG 3002  
Db 2959 GAAGAACTACTCTCTGTTGTCCGCTGATGTCAACACTTCTTAAATCAAAGGTATGAAG 3018  
Qy 3003 GAGCGGTGACAAACCACTTCACTGTACAAACGATCGGTCTCATCTGAGATGATAGGCGGTA 3062  
Db 3019 GACAGTAATGTGCACGTCTTCTGAAAAGAAATCAGTTTCCAGGAAATGGTTAGTGGGG 3078  
Qy 3063 AGGAGTACTAAACAGATGTTTCCAAACCACTTAAAGGGAATTTGTAACCTTCACTCAGG 3122  
Db 3079 CTGGCTATCAATCTGTGTCGAAGCGCTTAAAGGGAATTTTGAACCTTTCACAGT 3138  
Qy 3123 CTGATAAATTTGATTAAGAGAGAGGGCTATAGAATGTGAACACCGTTTCAAGATCC 3182  
Db 3139 CTGAAGAGGAGCCCTTCTCTCAAGGGGCTATGAGATGTCATCTACTGTACATGAGGTAC 3198  
Qy 3183 AAGGAGAAACCTTTGAAGATGTGTCGTGCTGATGAGTTGACGGGCACTCCACTGACTCTGA 3242  
Db 3199 AAGGTGAGACGTATGACAGCGTATCTGTTAGTTGAGCTTAAACCTTACCGCTGTATCTATCA 3258  
Qy 3243 TTTTCAAGTCTTCCCGCATGTTCTTAGTCGCTCTGACTAGACACACAAAGAGCTTCAAAAT 3302  
Db 3259 TCGCAAGAGACAGTCCCGCATGTTCTGTCTCGTTGTCAAGACACACAAATCCCTAAGT 3318  
Qy 3303 ATTACACCGTAGTGTAGATCCCTTTAGTACAGATAATAGTGTGTTCTTTTAAAGCT 3362  
Db 3319 ACTACACCGTTGTGATGGATCCCTTTAGTTAGTATCATTAGAGATTTAGAAACGGGTAGTA 3378  
Qy 3363 CCTTCTCTTTAGAAATGTATATGTTAGAACAGCAGGTAGTATAGCAATTAACAGATCGATG 3422

Db 3379 GTTACTTTATTAGACATGTCACAAAGTAGATGCAGGTACTCAATAGCAATTAACAGTCCGACT 3438  
Qy 3423 CAGTGTTCAAAGTCAATATCTCTTTGTGGCAACACCTTAATCAGGAGACTTTTCCAGATC 3482  
Db 3439 CTGTGTTTAAAAATTTCAATCTTTTGTAGCAGCTCCAAAGACTGGAGATATATCTGATA 3498  
Qy 3483 TACAGTTCATTAACGATGATGCTCCCTGGTAAATAGTACTATATCTTAACAAGATGATG 3542  
Db 3499 TGCATTTTACTATGATAGTGTCTTCTGGGAAACAGCAGTGTGTGAAACACTACGACG 3558  
Qy 3543 CTGTTACCATGAGTTTACGTGATAATAGTCTTAAATGTGAAGGATGTGTTCTTTGATTTTT 3602  
Db 3559 CTGTTACCATGAATTTGACTGACATTTCTTGAATGTCAAAGATGTGATATTAGATATGT 3618  
Qy 3603 CCAAAAGTATTCGATGCCAAAGAGAGGTGAACCAATGTCTAGAGCCAGTTTTCGCTACCG 3662  
Db 3619 CTAAGTCTGTAGTGTCTCCGAAAGATGTCAAACCAACTTTTAATACCAGTGTGTACGAACGG 3678  
Qy 3663 CGCGGAAACCGCCCAAGGGCTGCAGGACTACTCCGAAATCTGGTTGCAATGATTTAAAAGAA 3722  
Db 3679 CGCAGAAATGCCCTCCGACACTGGACTGTGGGAAATCTAGTTGCAATGATTTAAAAGAA 3738  
Qy 3723 ATTTCAACCGCACAGACCTGACGGGACGATTTGACATTTGAGAGCACCGCATCTGTTGTAG 3782  
Db 3739 ATTTTAAATCACGAGAGTTGTCCGAGTAGTTGATATTGAAAATACTGCATCTTTAGTGG 3798  
Qy 3783 TAGATAAGTTTTTGTAGTATTTTTTAAABAAAGAAAATACACAAAATATTGCTG 3842  
Db 3799 TAGATAAGTTTTTGTAGTATTTTTTAAAGGAAAAGAAAACCAAAACAAAATTTTTT 3858  
Qy 3843 GAGTGTATCGAAGAGTTCATATGATGAGATGGTTGAAAACAGGAAAGTACTATTGG 3902  
Db 3859 CACTGTTTAGTAGAGAGTCTCTCAATAGTGGATAGCAAGCAAGCAAGTCAATTTG 3918  
Qy 3903 ACAGCTTGGTAACTACAAATTTTACAGATCTGCGGGCCATTCGATACGTACAGCAACATGA 3962  
Db 3919 GTCAGTTGGCCGATTTTGACTTTTGTGGATCTTCCAGCCGTTGATCATGATCAGGCATATGA 3978  
Qy 3963 TCAAGCTCAACCAAAACAGAAATTTGACCTTTCAATTCAGAAATGAATACCTGCTCTGC 4022  
Db 3979 TTAAGCGCAACCGAAGCAGAAACTGGATCTGTCAATTCAGACAGAAATTCAGACGTTGC 4038  
Qy 4023 AAACAAATTTCTACCAATTCGAAAGCAGATCAACGGTATTTTGG---CCGGTTTCTCAGAGC 4079  
Db 4039 AAACGATTTGTATCATTTCAAAGAAATCAAACGCAATATTTGCTCTCTTTTCAGTGAGC 4098  
Qy 4080 TTACAGGTTGCTGCTCGAGGCAATTTGATTTCAAGAGTTTCTTTTCTTTTCTAGAGAAA 4139  
Db 4099 TTACAGGCAATTTACTTGAACGATTTGACTCAAGCAGATTTCTTTTCTTTTACGAGAAAGA 4158  
Qy 4140 CTCGAAACAGATTTCAAGAAATTTTCTCGGATCTCGACTCGCAGCTTCTATGGATGTGT 4199  
Db 4159 CACCGCTCAGATCGAAGATTTCTTCGGAGATCTAGACAGTCAATGTCCCAATGGACGTAC 4218  
Qy 4200 TAGAACTGATATTTCTAAGTATGATAAGTCAACAGAGTTTCAATGTCTGTAGAGT 4259  
Db 4219 TTGAGTTGAGTGTTCGAAAGTATGACAGTCCCAAAACGAGTTTCAATGTCTGTGTAGT 4278  
Qy 4260 ATGAATATGAAAAGATTTGGGTCTCAATGAGTTTTCGCCGAGTGTGGAAACAGAGGC 4319  
Db 4279 ACGAAATCGGAGAGACTGGGTCTGGAGATTTCTTTGCAAGGTGTGGAAACAGAGGC 4338  
Qy 4320 ACAGAAAAACAATTTGAAGGATTAATTTGCTGGAAATCAAGACATGCTCTGTGGTATCAAA 4379  
Db 4339 ATAGAAAAACCACTCTGAAGATTAACATGCTGGTATAAAAACGTTTATGGTACCAGA 4398  
Qy 4380 GGAAGACGGTGTAGTGTACTTTTCATCGGCAATCTGTTTATAATPAGCAGTGTGCTGG 4439  
Db 4399 GAAAGAGTGTGTATGTTTCAAACTTTTATCGGTAATACCGTCAATGCTTCTGCTGTCTG 4458  
Qy 4440 GTTTCATGTTTACCGATGGAAGAGGTATATAAAGGTGCTTTTGTGGAGACGATTCGTTT 4499  
Db 4459 CATCAATGCTCCGATGGAAAAATTTGATAAAGGAGCCTTCTCGGAGATGATAGTTGT 4518

4500 TGTATTTTCCAAAGGTTTGGATTTCCCTGACATTCAGTCAATGCTGCTAATCTCATGTGGA 4559  
4519 TGTACTTTTCCAAAGGTTTGGATTTCCCTGACATTCAGTCAATGCTGCTAATCTCATGTGGA 4578  
4560 ATTTTGAGGCCAAACTGTATAGAAAGAGTACGGTTACTTTTGTGGTAGATACATCATAC 4619  
4579 ATTTTGAGGCCAAACTGTATAGAAAGAGTACGGTTACTTTTGTGGTAGATACATCATAC 4638  
4620 ACCATGATAGAGGAGCAATAGTGTATATATGATCTTTTGAAGTGTATCTCCAACTTGGGG 4679  
4639 ATCAGCATAGAGGTGTCATAGTATATCTACGACCTTTTGAAGTGTATCTCCAACTTGGGG 4698  
4680 CAAAACATATCAGGATATATGATCATTAGAGGTTTAAAGGTTCTTTTGTGGTAGATCTTC 4739  
4699 CTAACACATCAGGATATGATCATTAGAGGTTTAAAGGTTCTTTTGTGGTAGATCTTC 4758  
4740 CTTGTTGCTCGGAACTGCTGCTTAGGCTTTCCGACGCTGAAAGCAGCTATCAAGGAGG 4799  
4759 CTGAGTCTTTTGAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTT 4815  
4800 TTTCAAAAACCGGATATGATGTTTGGTTTCTTTTAAATGTTTAAACAAATTTTGTGTTG 4859  
4816 TTTCAAAAACCGGATATGATGTTTGGTTTCTTTTAAATGTTTAAACAAATTTTGTGTTG 4875  
4860 ATAAATTTTATTTAGAACTTTGTTTAAATGTTTAAATGTTTAAATGTTTAAATGTTTAA 4919  
4876 ATAAAGTTTATTTAGAACTTTGTTTAAATGTTTAAATGTTTAAATGTTTAAATGTTTAA 4935  
4920 ATTAGCGAGTTTATTTAGAACTTTGTTTAAATGTTTAAATGTTTAAATGTTTAAATGTT 4979  
4936 ATTAATGAGTTTATTTAGAACTTTGTTTAAATGTTTAAATGTTTAAATGTTTAAATGTT 4995  
4980 GTCAAGAGTTTATTTAGAACTTTGTTTAAATGTTTAAATGTTTAAATGTTTAAATGTTT 5039  
4996 GTAAAGAGTTTATTTAGAACTTTGTTTAAATGTTTAAATGTTTAAATGTTTAAATGTT 5055  
5040 TCTGATGTAGATTTATTTAGAACTTTGTTTAAATGTTTAAATGTTTAAATGTTTAAATG 5099  
5056 TCTGAAATTAATCTTTTAAAGGTTTAAATGTTTAAATGTTTAAATGTTTAAATGTTTAA 5115  
5100 GATTTGTTAGTTTCTGCGGAGTGGATCTCCCGATTAATCTCCGATTAATCTCCGATTA 5159  
5116 GGTCTTGTGTTGTTGCGGAGTGGATCTCCCGATTAATCTCCGATTAATCTCCGATTA 5175  
5160 TGTATTTAGTAAGAGATGAAAGAGTAAAGAGCAACGCTGGTGGCTATCAAGCC 5219  
5176 TGATGTTGCAAGAGATGAAAGAGTAAAGAGCAACGCTGGTGGCTATCAAGCC 5235  
5220 CTTGCTTGCAAAAGATTTTCTTTTAAAGTAAATCTTAAATGTTTAAATGTTTAAATGTT 5279  
5236 GCTGCTGCTAAAAGCGGTTTCAAGTTTAAAGTAAATCTTAAATGTTTAAATGTTTAAAT 5295  
5280 GATGCTGAGAGCAACCGCTGGCAAGTTTAAAGTAAATCTTAAATGTTTAAATGTTTAA 5339  
5296 GATGCGAAAAGAGCAATATGCGAGGTTTAAAGTAAATCTTAAATGTTTAAATGTTTAA 5355  
5340 GGATCTGCTCTTTATCTTTTGGAGTTTCTTTTAAATGTTTAAATGTTTAAATGTTTAA 5399  
5356 GGCTACTGCTCTTTATCTTTTGGAGTTTCTTTTAAATGTTTAAATGTTTAAATGTTTAA 5415  
5400 AGAAAAGTTTGGAGGAACGTTATTTTGGAGTTGTCAGACAGCGCTCGCCAAATTTGAAC 5459  
5416 AATTTGTTTGGAGGAAGTAAACGAGTGTGAACGATGAGGAGCCATGGAACCTTCA 5475  
5460 GAAAAGTTTGGAGGAGTTCTGAGATGAAGATCAATAGCTGTGAACTCGAAAAGGTT 5519  
5476 GAAAAGTTTGGAGGAGTTCTGAGATGAAGATCAATAGCTGTGAACTCGAAAAGGTT 5535  
5520 CCGGAAAACAAAAGAGTGTAGGTAATTAATCTTAAATGTTTAAATGTTTAAATGTTTAA 5561  
5536 CGAACCAATCTCTCAAAAAGAGGTCGGAAGAAATTAATTAATTTAGGTAAAGGCGGTTCA 5595

5562 ---AAGAAAATAAATAACAGTGGTAAGAGGGTTTAAAAATTGAGGAAATTGAGGATAAT 5618  
5596 GCGGGAAGGCTTAACCAAAAAGTTTGTAGTAAGATTGAAAAGAGTTTGTATATTTGATT 5655  
5619 GTAAGTGTAGCAGAGTCTATCGCGTCAATCGAGTACGTTTTTAATCAATATCCCTTATCAA 5678  
5656 GAAGATGAAGCCGAGAGCGTGGTTCGGATTCTGATTCTGATTAAATATGCTTACCCAA 5715  
5679 TCAACTCTCGAGCCAAATTTGTTTACTTATCTTCGCTTACGAGATCCCTGTGAGCGTGA 5738  
5716 TCACTCTTCCATCGCAATTTGTTTGTGTTTGTCTGTATGGGCTGACCCCTATAGAATTGT 5775  
5739 TCAATCTGTGTACAAATGCAATTCGGGTAAACCAAGTTTCAACGCAACAGCTAGACAAACAG 5798  
5776 TAAACGTTTGTACAAATTCGTTAGGTAAACCAAGTTTCAACAGCAAGCAAGCAACTACTG 5835  
5799 TCAACAGCAATTTGCGGATGCTCGGAAACCTGTGCTAGTATGACAGTGCAGATTTCCCTG 5858  
5836 TCAACAGCAATTTGCGGATGCTCGGAAACCTGTGCTAGTATGACAGTGCAGATTTCCCTG 5895  
5859 CATCGGATTTCTAGTGTATAGATATAATTCGAGCGCTTGTATCCGTTGATCAGCGCGTTAT 5918  
5896 GCGATGTTTATAGGTTGTATAGGTACAATGCACTTTTGTAGATCTTTTAAATTACTGCGTGC 5955  
5919 TAAATAGCTTTGATACCTAGAAATAGATAATAGAGTTTGTATATCAACCCGCAACCAATA 5978  
5956 TGGGAGCTTTGATACCTAGAAATAGATAATAGAGTTTGTATATCAACCCGCAACCAATA 6015  
5979 CTACTGAAATTCGTTTAAACGAGCTCAGAGGTTAGACGATGCTACTGTAGCTATTAAGGCTT 6038  
6016 CAGCTGAAAGTTAGATGCTTACCCGAGGTTAGACGCTACGTTGCAATTCGGTCTG 6075  
6039 CAATCAATTAATTTGGCTTAATGAATCGTTGCTGGAACCTGGCATGTTTCAATCAAGCAGCT 6098  
6076 CTATAAATAATTTAGTTTAAATGAATAGTAAAGGTTAGTACGCTGTACAAATCAGAATACTT 6135  
6099 TTGAGACTGCTAGTGAATGCTGCAACCAACTCGGCTACTTGTAGCTATTTGTTGTAG 6158  
6136 TTGAAAGTATGCTGCGGTTGGTCTGACCTCTGACCTGCTATCTTAAATGATAGGTTGCT 6195  
6159 ATTTCTCAATAAATAAGTCGCTGAAGACTTAAATAATCAGGTTGCTGATACCAAAATCAGC 6218  
6196 GAATATATAA-----GTTTGTGTTTCTTAAACACACGCTGCTACGATAACGTAC 6247  
6219 AGTGGTTGTTGCTGCTTAAATATTAACGATTCATATCTGATCTCAACAGATTAAACCA 6278  
6248 AGTGGTTTCCCTCCCACTTAAATCGAAGGTTAGTGTCTTGGAGCGCGGAGTAAACATA 6307  
6279 TGTGATGTTGTATATCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 6338  
6308 TATGGT-TCATATATGTTCCGTAGGCACTGTAAGGAGGAGGATTCGAATTTCCCGCGNA 6366  
6339 CCGCGGTTAGCGGCCCA 6355  
6367 CCGCGGTTAGCGGCCCA 6383

## RESULT 10

AF155507  
LOCUS AF155507  
DEFINITION Tobacco mosaic virus, complete genome.  
ACCESSION AF155507  
VERSION AF155507.2 GI:5809676  
KEYWORDS  
SOURCE Tobacco mosaic virus  
ORGANISM Tobacco mosaic virus  
REFERENCE 1 (bases 1 to 6384)  
AUTHORS Bingheng, Q., Gong, Y., Xiuhua, Z. and Po, T.  
TITLE Complete nucleotide sequence of attenuated tomato mosaic virus K  
JOURNAL genome  
Patent: China (00100211.2) 14-JAN-2000;  
CAS, Institute of Microbiology, Beijing, China

AF155507 6384 bp RNA linear VRL 02-MAR-2000

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

2 (bases 1 to 6384)  
Sheng, Q.B. and Gong, Y.  
Direct Submission  
Submitted (01-JUN-1999) Molecular Virology and Genetic Engineering,  
Institute of Microbiology, Chinese Academy of Sciences,  
Zhongguancun Beiyitiao No.13, Haidian District, Beijing 100080,  
People's Republic of China

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

3 (bases 1 to 6384)  
Sheng, Q.B. and Gong, Y.  
Direct Submission  
Submitted (31-AUG-1999) Molecular Virology and Genetic Engineering,  
Institute of Microbiology, Chinese Academy of Sciences,  
Zhongguancun Beiyitiao No.13, Haidian District, Beijing 100080,  
People's Republic of China  
Sequence update by submitter  
On Aug 31, 1999 this sequence version replaced gi:5524641.

REMARK  
COMMENT  
FEATURES  
source

1. .6384  
/location/Qualifiers  
/organism="Tobacco mosaic virus"  
/mol\_type="genomic RNA"  
/db\_xref="taxon:12242"  
/note="attenuated mutant strain originally derived from  
nitrite-treated tomato strain"  
72. .4922  
/gene="replicase"  
72. .4922  
/gene="replicase"  
/function="genomic replication"  
/note="read-through protein"  
/codon\_start=1  
/transl\_except=(pos:2670. .2672, aa:OTHER)  
/transl\_except=(pos:3420. .3422, aa:OTHER)  
/product="183 kDa protein"  
/protein\_id="AAD44327.1"  
/db\_xref="GI:5524644"

gene  
CDS

/translation="MAYQTATSSALLETVRGNTLVNDLAKRRLVDAADVDFNARDR  
RPKNFSGVSEEQTLIATKAYPEFQITFYNTQNAVHSLAGGRLSELEYLMQIIPYG  
SLTYDIGNFASHFKGRAYVHCMPNLDVDMRHEGQKDSIELYLSRLERGNKHPV  
NFOKEAFDRAEMNEVCHDTQTCTHSEOECYTRGVYATALHSIYDIPADEFGAALL  
RKNVHCYAAFHSENLLEDHVNDEINACFORQDRLTFSESTLNSHSYSN  
ILKYCKTTPASNRVYKKEFLVTRVNTWFKFSRIDTFLLYKGVAHKGVDSQEPFK  
TMEDAHYKTLAMNSERILLEDSSVNTWFFPMRDMVLPFLDISLETKRTRKEV  
LVSKDFVTLNHIRTQKALTYSNLSESVISRVVINGVTARSEVDKSLLOS  
LSMTFFLTKLAVLKDDLLISKFALGKTVSQHWDEISLAFGNAPFSIKERLINRKL  
IRITENALEIRVDPVLTVDHRLVSEYKMSVDMPLDIRKMBEETEMYNALSELV  
KNSDKDGVFSQMSQSLVDPMATAKVIIVAMNSESGLTLTPEQTEANVALAODS  
EKASDGLAVTPRDEEPSIKGSMARGELQAGLSDGVPSESTRSEESLQEFHMA  
TASLIHKQMCISIVYTGPLKVOQKNFIDSLVASLAAVSNLVKILKDTAAIDLETQ  
KFGVLVDSKRWLVKPSADHGWVETHARKYRVALLEHDEFGIITCDNWRVAVSS  
ESVYSDMAKRLTLRLKDGEPHVSAAKVLVDGPGCGTKTEILSRVNFEDLLIV  
PGRAAEMIRXRANASGIIIVATKDNVTVDSFLMNYGKRAKCOFKRFLIDEGMLHTG  
CYNPLVMSLCDIAYVYGDQQIIPYINRVTFGPYPAPFAKLEVDETRTTLRCPAD  
VTPLNQRVYEGHVMCTSSSEKSVSQEMVSGAASINPVPKLGKILFTQSDKEALLS  
RGYADVHTVHEVQETVADVSLVRLTPTPVSIIARDSPHVLVSLSRHSLKTYTVVM  
DPLVSIIRDLRVSSYLLDMYKVDAGTQ"

72. .2672  
/gene="replicase"  
/function="genomic replication, virus attenuation and  
specific protection against wild challenging viruses"  
/note="contains opal mutation"  
/codon\_start=1  
/product="98.5 kDa protein"  
/protein\_id="AAD44325.1"  
/db\_xref="GI:5524642"

/translation="MAYQTATSSALLETVRGNTLVNDLAKRRLVDAADVDFNARDR  
RPKNFSGVSEEQTLIATKAYPEFQITFYNTQNAVHSLAGGRLSELEYLMQIIPYG  
SLTYDIGNFASHFKGRAYVHCMPNLDVDMRHEGQKDSIELYLSRLERGNKHPV  
NFOKEAFDRAEMNEVCHDTQTCTHSEOECYTRGVYATALHSIYDIPADEFGAALL  
RKNVHCYAAFHSENLLEDHVNDEINACFORQDRLTFSESTLNSHSYSN  
ILKYCKTTPASNRVYKKEFLVTRVNTWFKFSRIDTFLLYKGVAHKGVDSQEPFK  
TMEDAHYKTLAMNSERILLEDSSVNTWFFPMRDMVLPFLDISLETKRTRKEV  
LVSKDFVTLNHIRTQKALTYSNLSESVISRVVINGVTARSEVDKSLLOS  
LSMTFFLTKLAVLKDDLLISKFALGKTVSQHWDEISLAFGNAPFSIKERLINRKL  
IRITENALEIRVDPVLTVDHRLVSEYKMSVDMPLDIRKMBEETEMYNALSELV  
KNSDKDGVFSQMSQSLVDPMATAKVIIVAMNSESGLTLTPEQTEANVALAODS  
EKASDGLAVTPRDEEPSIKGSMARGELQAGLSDGVPSESTRSEESLQEFHMA  
TASLIHKQMCISIVYTGPLKVOQKNFIDSLVASLAAVSNLVKILKDTAAIDLETQ  
KFGVLVDSKRWLVKPSADHGWVETHARKYRVALLEHDEFGIITCDNWRVAVSS  
ESVYSDMAKRLTLRLKDGEPHVSAAKVLVDGPGCGTKTEILSRVNFEDLLIV  
PGRAAEMIRXRANASGIIIVATKDNVTVDSFLMNYGKRAKCOFKRFLIDEGMLHTG  
CYNPLVMSLCDIAYVYGDQQIIPYINRVTFGPYPAPFAKLEVDETRTTLRCPAD  
VTPLNQRVYEGHVMCTSSSEKSVSQEMVSGAASINPVPKLGKILFTQSDKEALLS  
RGYADVHTVHEVQETVADVSLVRLTPTPVSIIARDSPHVLVSLSRHSLKTYTVVM  
DPLVSIIRDLRVSSYLLDMYKVDAGTQ"

72. .5634  
/gene="Mp"  
/function="virus movement in plants; involved in local  
lesion development and deformity of tobacco leaves"  
/note="movement protein; contains ochre mutation"  
/codon\_start=1  
/product="26.8 kDa protein"  
/protein\_id="AAD44328.1"  
/db\_xref="GI:5524645"

/translation="MALVVKGVKYNINEFIDLSKSEKLLPSMTFVKVSMVSKVDKIMV  
HENESLSEVNLKGVKLEGGYVCLVGLVSGEWNLPDNCRGVSVCMVDMKMRABE  
ATIGSYTAAAKRQFQKVPVNPYITTKDAEKNIWOVLVNIKNVMSAGYCPILSLFV  
SVCIYKNNIKLGLREKVTSDNGGPMELSEEVDFEWMENVPMSRLAKFRTKSKSG  
PKNNNLKSGRSGRPKPSFDEV"

5703. .6182  
/gene="Cp"  
5703. .6182  
/gene="Cp"  
/function="virus assembly"  
/note="coat protein"  
/codon\_start=1  
/product="17.7 kDa protein"  
/protein\_id="AAD44329.1"  
/db\_xref="GI:5524646"

/translation="MSYSITSPSQFVFLSSWADPIELLNVCTNSLGNQFQTOAARTT  
VQOOFSEVWKPFQSTVRFPGDVKYVRYRNAVLDPDLITALLGADFTNRRIIEVENQOS

gene  
CDS

72. .3422  
/gene="replicase"  
/function="genomic replication"  
/note="read-through protein"  
/codon\_start=1  
/transl\_except=(pos:2670. .2672, aa:OTHER)  
/product="126 kDa protein"  
/protein\_id="AAD44326.1"  
/db\_xref="GI:5524643"



PTTAETLDATRRVDDATVAIRSAINLVNELVRGTGLYNQNTFESMSGLVWTSAPAS"

## ORIGIN

Query Match 41.3%; Score 2621.6; DB 14; Length 6384;  
Best Local Similarity 64.2%; Pred. No. 0;  
Matches 4096; Conservative 0; Mismatches 2224; Indels 60; Gaps 8;

QY	18	CGACAACAACTTAAACAAACAAACAAACATATTACAAACAAACAAACAAACAAATGGCAC	77
DB	23	CAACAACAAACAAACAAACAAACATTTACATTTTACATTC-----TACAACTACAAATGGCAT	78
QY	78	ACATACAACTCTATAATAGCAACGCTCTCTCAAGAGCGTGAGTGGTGAACAACTCTCG	137
DB	79	ACACAACAAACGACCATCGTCTCGCTTGTGTGAGCCGTCGAGGTACAACTCTGG	138
QY	138	TTAATGACCTTGAAGAAGCGCATGTACGATACGCGCGTGGAGAAATTTAACCCCGCG	197
DB	139	TCAACGATCTTGCACAAAGCGGCTATATGACGACGCGTGCATGAATTTAATGCTAGGG	198
QY	198	ACGTAGACCAAGGTCACTTTTCCAAACTATTAGCGAAGCAACGCTTCTAGTCT	257
DB	199	ACGCGAGGCTTAAAGTCAATTTTCCAAAGTAGTAAGCGAAGAACAGACGCTTATTGCAA	258
QY	258	CCAAACGCTACCGGAGTCCAGATTACCTTTTATTAATCTCAAAATGCCGTACACAGTT	317
DB	259	CCAAAGCTACCGAGAAATCCAAATTTACATTTACACACGAGAAATGCTGTGCATCCC	318
QY	318	TGGCTGGAGTTTGGAGGCAATAGAAATGGAATATCTGATGCTACAAATTTCCCTATGGAT	377
DB	319	TTGCGAGCGTCTCCGATCATTAGAAATTTGAATATCTGATGATGCAAAATTTCCCTACGGAT	378
QY	378	CGCGGACATATGATATAGTGGGAACTTTGCGAGCACTTTGTTTCAAGGCGAGGATACG	437
DB	379	CATTGACATATGATATCGGAGGTAATTTTGCATCTCATCTGTTTCAAGGCGGAGCATACG	438
QY	438	TGCATTGCTGTATGCCCAATCTCGACATACAGATATAATGAGCGACGAGGACAAAGG	497
DB	439	TTCACTGCTGTATGCCCAATCTCGATGTCGCGACATAATGCGGCGAGGCGGCAAAAGG	498
QY	498	ACTCAATTTGAGATGTATTTGTCAGAAATTTCTGGTTCTTAAAGGTAATTTCTGAGTTTC	557
DB	499	ACAGTATTGAACATATACCTTTCTAGGCTCGAGAGGGGCAACAACATGTCCTCAAACTCC	558
QY	558	AAAGGAGGCTTTTAAACAGATATGCGAAGCTCTCCAAAGAGTCTGCTGCTCTTAAACTTT	617
DB	559	AAAAGGAAAGCTTTTCCAGACATACGCTGAAATGCGCAACGAGTAGTCTGTCAAGATATT	618
QY	618	TTCAGGATTTGCAATATACCCCGCAGAGATAGTGTAGAGATACGCTGTTGCTCTGC	677
DB	619	TCCAAAGCTGTAGGCACTTCAAGAAATGTTTACACGGGAGAGTGTATGCTTATGCTTTGC	678
QY	678	ACAGTTTGTATGATATTTCTGTGATGTTTGGAGCTGGGTTTAAATATCTTAAAGATATAC	737
DB	679	ATAGTATATAGATATACCTGCGGACGAGTTGCGGCGGCACTGCTGAGAAAGAAATGTAC	738
QY	738	ATGATATGATGAGCTTCCATTTTGGCAGAGCAATTTATTTACAGCAGACGAGGTTA	797
DB	739	ATGATATGATGAGCTTCCACTTTTCCGAGAAATTTACTTCTCGAAGATTTCAACAGTCA	798
QY	798	CGCTTAATGAATAGGCGCACTTTTCAAGAGAGAGTGTATGTTTCTTTTCTTTTCTTTG	857
DB	799	ACCTCGACGAGATCAATGCAATGTTTCCAAAGAGATGAGACAGGTGCACTTTTCTTTT	858
QY	858	CTGATGAAGTACTTTTAAATTTATAGTCTAAATACAAATAATATCTTGCATTTATGTAGTTA	917
DB	859	CATCTGAGATCTCTTAAATTTATAGTCTATAGTATTCTTAAATTTTAAAGTATGTTTGA	918
QY	918	AATCTTACTTTCTGCTCTCTAGTAGAAATGTTTACTTTTAAAGAAATTTTATGCTACAGG	977
DB	919	AAACTTACTTTCCAGCTCTTAATAGAGAGGTTTACATGAAGGAGTTTATAGTAACTAGAG	978
QY	978	TTAATACTTGGTTTGTAAATTTTACAAAGTAGATCTATATCTCTGTAACAGAGTTTA	1037

DB	979	TTAATACCTGGTTTGTAAATTTCTAGAAATAGATACTTTCTTTATTGTACAAAGGTGTAG	1038
QY	1038	GACAACTAGGGTGTAGTAGTGTATGATTCATATGAGCGGATGGAAGAGCGCTTTGCTTACA	1097
DB	1039	CGCATAAAGGGTGTAGTAGTGTAGTGTATTTTCAAGAGCTATGGAAGAGCGATGCGACTACA	1098
QY	1098	AGAAAACCTTGGCCATGTTTCAACACATGAAGAGCAATCTTTTACAGACACCGCTTCGGTTA	1157
DB	1099	AAAAGACTCTTGGGATGTACAACGTGAGAGAAATCTTTGTAGAGGATTTCTTCAATCAGTTA	1158
QY	1158	ACTTTTGGTTTCCCTAAAGATGAAGGACATGGTGTATAGTACCGCTGTTTGGAGGTTCTATTA	1217
DB	1159	ATTACTGGTTTCCAAAATAGGGATATGGTGTATGTTCCACTATTGACATATCTCTCG	1218
QY	1218	CMAGCAAAAGATGACAAAGAGTGAAGTCAATGTTTAATCGTGTACTGTTTACACAGTGC	1277
DB	1219	AGACTAATAAAGAAACACGCAAGAGGCTTCTAGTTTCAAGAGCACTTTGTTTATACAGTGT	1278
QY	1278	TTAATCATATCAGAACATATCAAGCCAAAGCGTTAACTTACCAGAAAGCTATTTATCTTTTCG	1337
DB	1279	TAAATCACATTTCTGATCCAGGCCAAAGCGCTTACTTACTCAAGCTGTATCTTTTCG	1338
QY	1338	TGAGTCTTATAAGATCCCGCTGATTAATCAATGGTGTACTGTAGTCTTGAATGGGATG	1397
DB	1339	TCGAATCAATTCGTTGAGAGTGTATTAACGCGGTACTGTAGTCTGAGTGGGATG	1398
QY	1398	TAGATAAGCAATCTTCAACCTTGTCAATGACTTTCTTCTTTCGAGACTAAGCTGGCTG	1457
DB	1399	TCGATAAATCAATTTATPACAGTCTTCTCGATGACGTTCTTCTTACACACCAAGCTTGGCG	1458
QY	1458	CGCTTCAAGACGATATAGTAATGGGAAAGTTTCGGTGTCTGGATAAGACCACCTTCTGAAC	1517
DB	1459	TTCTGAAAGACGATCTTTTGAITAGCAAGTTTGCACTTGGACCAAAAACCTGTCTCAACAC	1518
QY	1518	TTATTTGGGATGAGGTGGGCAAAATTTTGGAAACGTTTTTCCCACTATCAAAAGAGAGAT	1577
DB	1519	ATGTGTGGGATGAGATTTCCCTAGCTTTCCGCAATGCTTTTCCCTCGATCAAGAAAGAT	1578
QY	1578	TGAGTACGAGGAAATTTCTGATGTAAAGTGAAGTGTCTTGAAGATCAAGATCCAGATC	1637
DB	1579	TGATAAACCGGAACTGTACAGAAATTAACGAGAAATGCGTTAGAGATCAAGGTGCGCCATC	1638
QY	1638	TGTTATCTCACATGGAAGACAGGTTCTGATGCTCAATACCAAGTCTGAGGAGTTACCGC	1697
DB	1639	TTTATGTCTCTTCCATGATAGGTTAGTTTCTGAGTACAAAATGTCTAGTGGACATGCCGG	1698
QY	1698	ATCTAGATATCAAGAAAGCACTTAGAAGAGCTGAGCAAAATGTACGCGCTTATCAGAAAT	1757
DB	1699	TGCTAGACATTAGGAAAGAGATCGAAGAACTGAGGAAATGTACAAATGCACTGTCCGAAC	1758
QY	1758	TATCTATCTTAAAGGTGTGATAATTTGATATCGGAAAGTTCAAGAGATGTGCAAGG	1817
DB	1759	TGTTCTACTTAAATAATTCAGACAACTGATGTGACGTTTTTTTCCAGATGTGCCAAT	1818
QY	1818	CTTTAGATGTAGTCTCTGATGTGGCAGCAGAGTAACTGTTGAGTGGCCGAGAAATAGAA	1877
DB	1819	CTTTAGAGTGTGATCCAAATGATGTCAGCAAAAGGTAATAGTAGCAGTTATGAGCAACGAG	1878
QY	1878	GCGGTTTAACTCTTACTTTTGAAGCAACCGAGGAGATGTGGGCTTAAGGCTCTTAAA -	1936
DB	1879	GTGGTCTTACTCTCACTTTGACGCCCAACCGAGCTAATGTTCGCTAGCTAGCTGCAAG	1938
QY	1937	-----AGCACGGCTCTGAGCGCGTGTATGTCTTGAACCGCATCCGAAAGAGGTGAACG	1991
DB	1939	ATTCTGAAAGGCTTCTGATGGGCGTGTGTAGTTACTCCCAAGAGATGTTTGAAGAACCGT	1998
QY	1992	TAAATAAATTTTCTATTTGCTGAGAAAGGAGATTTGCTGTGTGCGAAGATCATGGTT	2051
DB	1999	CCATNAAGGGTTCGATGGGCGCGTGGTGAAGTTACAAATGGCCGATATCTTGGGACGCTC	2058
QY	2052	TGACGAATGCTAACTTTAGAGCACAGAGTGTGGAGTCCCTCAACGATTTCTCAATAGGCTT	2111
DB	2059	CTGAATCTTCTATACACTAGGAGCGAGGAGATTGAGTCTCTCGAGCAGTTTCTATATGCAA	2118

Qy 2112 GCCTGGATGCTGATTAACAAGCAATGGCATCGGTGTCTACACTGGCTCACTCAAAG 2171  
Db |||||  
Qy 2119 CAGCTAGTTCGTTAAATTCATAAGCAGATGTGTCGATCGTGTACACGGGCCCTCTTAAAG 2178  
Db |||||  
Qy 2172 TTCAACAATGAAGAACTATGTGACAGATTTGGCAGCTTGGTGTCCGCCACTGTATCAA 2231  
Db |||||  
Qy 2179 TTCAACAATGAAGAACTTATATAGACAGCTGTGTAGCTCGCTCTCTGCTGCGGTGTCGA 2238  
Db |||||  
Qy 2232 ATCTATGCAAGTCACTAAAGGATGAAGTCCGGTATGATTTCTGATTTCCAGGAGAGAAAGTGG 2291  
Db |||||  
Qy 2239 ATCTAGTGAAGATCCTTAAAGATACACCGCGATTGACCTTGAACCTCGTCAAAAGTTCG 2298  
Db |||||  
Qy 2292 GTGTTTGGGATGTCATTTTGAAAGTGGTCTCTCAAACTCGCGGCCAAAGGTCATTCAT 2351  
Db |||||  
Qy 2299 GAGTTCCTGGATGTGCTTCGAAAAAGTGGCTAGTTAAACCATCGCAAGACCATGCAT 2358  
Db |||||  
Qy 2352 GGGGAGTTGCTGGATTAACAGGGGAAATGTTTACTGCACTCTATCTATTAAGAAGGAG 2411  
Db |||||  
Qy 2359 GGGGGGTTGTGAGACTCATCGGAGAAATATACGCGTCGCAATTACTGGAGCACGATGAAT 2418  
Db |||||  
Qy 2412 ATAGAAATGGTCACTGAGAGGACTGGAGGGGTGGCTGTATCATCTGTATACATATGAT 2471  
Db |||||  
Qy 2419 TTGCAATTTTACGTCGATTAACCTGGCGACGGGTGGCTGTGAGTTCTGAGTCGGTAGTAT 2478  
Db |||||  
Qy 2472 ATTCTGATATTGCAAAAGCTCCAAATCTGAGGAAACAAATGAGAGCGGTGAACCCCAAG 2531  
Db |||||  
Qy 2479 ATTCTGATATGGCTAAACTCAGGACTCTGAGAGATTTGCTCAAAAGATGGAGACCAACAG 2538  
Db |||||  
Qy 2532 AACCTACTGCAAGATGGTACTTGTGATGGGTGGCTGTGTTGTGGAAAGTACAAAGGAG 2591  
Db |||||  
Qy 2539 TTAGTTTCAAGAAAGTGGTGTGTTGGTGGATGGCTTCCAGGGTGGCGGAACAAAGGAAA 2598  
Db |||||  
Qy 2592 ATTTTGAAGATTTGATCTTGATGAGGATTTGATCTTGGTTCTTGGAAAAACAAGTCTGTG 2651  
Db |||||  
Qy 2599 TTCTTTGAGAGTTAAATTCGAGAGAACCTTAATCTTGTCCCTGGTGTCTCAAGTGGCG 2658  
Db |||||  
Qy 2652 CTATGATCAGAAGAGGGCTAAATTCATCTGCACTGATTAAGAGCCACAAATGACAAATGTA 2711  
Db |||||  
Qy 2659 AGATGATCAGATGAAGAGCTAAATGCTGGCGCATTAATAGTGGCTTACAAAGGATAATGTGC 2718  
Db |||||  
Qy 2712 GAAAGTATGATCTTCTATGATCCAAACCGCGATCAC-----ACAAAGGCG 2762  
Db |||||  
Qy 2719 GCACCGTCGATTCATTTTATGATGAATTAACGGGAAAGGGCAGCGTGCAGTTCAAAAGAT 2778  
Db |||||  
Qy 2763 TTTTATTTGATGAAGGGTGTGATGCTGCACACCGGTTGTGTAACTTCCCTGGTGTCTTATCT 2822  
Db |||||  
Qy 2779 TGTTTATAGACGAGGTTTGTGCTGCTATCTGCTGTGTGMAATTTTGTGTTGMAATGT 2838  
Db |||||  
Qy 2823 CTGTTTGGGACATCGCATATATTTACGGAGATACACAGCAGATTCCTTTTCAATTAACAGAG 2882  
Db |||||  
Qy 2839 CTCTGTGCGATATTGCAATATGTTTATGGAGACACCCAAACAGATTCGCTACATCAACAGAG 2898  
Db |||||  
Qy 2883 TTCAGAAATTTCCGCTATCCCAACATTTTGGAGAGCTGCAAGTGGATGAAGTTGAGATGA 2942  
Db |||||  
Qy 2899 TAACTGTTTCCCGTACCTTCGACACACTTTCGAAATTTGGAGGTGCGAGAGAGTGGAAACAA 2958  
Db |||||  
Qy 2943 GGAGGACACACTGAGATGCCAGGATGATGTAATTTTTTCTTACAAATCGAAGTACGAAG 3002  
Db |||||  
Qy 2959 GAAGAACTACTCTCTGTTGTGCGGCTGATGTCAACACTTCTTAAATCAAAGGTATGAAG 3018  
Db |||||  
Qy 3003 GAGGGGTGACAAACACTTCAACTGTGAACAAGTCGGTCTCATCTGAGATGATAGGGGTA 3062  
Db |||||  
Qy 3019 GACAGTAAATGTGACGCTCTTCTGAAAAGAAATCAGTTTCCAGGAAATGTTAGTGGGG 3078  
Db |||||  
Qy 3063 AGGAGTACTTAAACAGAGTTTCCAAACCACTTAAAGGGAAATTTGTAACCTTCACTCAGG 3122  
Db |||||  
Qy 3079 CTGGCTCTATCAATCTCTGTGTCGAAGCCGCTTAAAGGAAATTTTTCAGTTTTCACAGT 3138  
Db |||||  
Qy 3123 CTGATAAATTTGAGTATAGAGAGAGGGCTATAGAAATGTGAACACCGGTTTCATGAGATCC 3182  
Db |||||  
Qy 3139 CTGACAAAGGAGGCCCTTCTCTCAAGGGGCTACCGCAGATGTCCATCTACTGTATCAGGTAC 3198  
Db |||||

Qy 3183 AAGGAGAAACCTTTGAAGATGTGTCGCTGGTCAAGTTGACGGCAACTCCACTGACTCTGA 3242  
Db |||||  
Qy 3199 AAGGTGAGACTTATGACAGCGTATCGTTAGTTTCGACTAAACACCTACGCTGTATCTATCA 3258  
Db |||||  
Qy 3243 TTTTCCAAGTCTTCCCGCATGTTCTAGTCTGCTCTGACTAGACACACAAAGAGCTTCAAT 3302  
Db |||||  
Qy 3259 TCCAAAGAGACAGTCCGCGATGTTCTGGTCTCGTGTCAAGACACACAAATCCCTAAAGT 3318  
Db |||||  
Qy 3303 ATTACACCGTAGTGTGTTAGATCCCTTTAGTACAGATAATTAAGTGTGTCTTCTTTAAAGCT 3362  
Db |||||  
Qy 3319 ACTACACCGTTGTGATGGATCCCTTTAGTTAGTATCATTAGAGATTTAGAACGGGTAGTA 3378  
Db |||||  
Qy 3363 CTTTCTTTTGAAGATGATATATGTTAGAGCAGGTAGTAGATAGCAATTTACAGATGGATG 3422  
Db |||||  
Qy 3379 GTTACTTATTTAGACATGTACAAAGTAGATGACAGTACTCAATAGCAATTTACAGGTGCACT 3438  
Db |||||  
Qy 3423 CAGTGTTCAAAGTGCATTAATCTCTTTGTGCAACACCTAAATCAGAGAGCTTTCCAGATC 3482  
Db |||||  
Qy 3439 CTGTGTTTAAAAATTTCAATCTTTTGTAGCAGCTCCAAAGACTGGAGATATATCTGATA 3498  
Db |||||  
Qy 3483 TACAGTCTTATTAACGATGTATGCTCCCTGGTAAATAGTACTATCTAAACAAGTATGATG 3542  
Db |||||  
Qy 3499 TGCAAATTTTACTATGATAGTGTCTTCTGGGAACAGCAGCTGTGTGAACAACATACGACG 3558  
Db |||||  
Qy 3543 CTGTTACCATGAGGTTTACGTTAATAGTACTTAAATGTGAAGGATGTGTTCTTTGATTTTT 3602  
Db |||||  
Qy 3559 CTGTTACCATGAAATTTGACTGCATTTCTGAAATGTCAAAAGATTTGCATATTAGATATGT 3618  
Db |||||  
Qy 3603 CCAAAAGTATTCGATGCGCAAGGAGGTGNAACCATGCTTAGAGCCAGTTTTCGCTACCG 3662  
Db |||||  
Qy 3619 CTAAGTCTGTAGTGTCTCCGAAAGATGTCAAAACCAACTTTAATACCAGTGGTACGAACCG 3678  
Db |||||  
Qy 3663 CGCGCAACCGCCCAAGGGCTGCAGGACTACTCGAAAAATCTGGTTGCAATGATTTAAAGAA 3722  
Db |||||  
Qy 3679 CGGCAAGATTCCTCCGACACTGGACTGTGTGAAATCTAGTTGCGATGATTAAGAA 3738  
Db |||||  
Qy 3723 ATTTCAACGACACAGACTGACGCGGACGATTTGACATTTGAGAGACCGCATCTGTTGTAG 3782  
Db |||||  
Qy 3739 ATTTTAAATTCACAGAGTTGTCCGAGTAGTTGATATTGAAATACTGCACTCTTTAGTGG 3798  
Db |||||  
Qy 3783 TAGATAAGTTTTTTGATAGCTATTTTAAAGAAAGAAATAACACAAAAATATTGCTG 3842  
Db |||||  
Qy 3799 TAGATAAGTTTTTTGATAGTTTATTTACTTAAGNAAAAAGAAACCAACAAAAATTTTT 3858  
Db |||||  
Qy 3843 GAGTGTATCAGAGAGTTCAATGATCAGATGTTGGAAAAACAGGAAAGAGTACTATTGG 3902  
Db |||||  
Qy 3859 CACTGTTTGTAGTAGAGATCTCTCAATAGTGGATAGTACAAAGCAAGAACCAATTCG 3918  
Db |||||  
Qy 3903 ACGACTTGGCTAATCAATTTTACAGATCTGCCGCCATCGATCAGTACAGACACATGA 3962  
Db |||||  
Qy 3919 GTCAGATGGCCGATTTTGTGATTTTGTGGATCTTCCAGCCGTTGATCAGTACAGGCATATGA 3978  
Db |||||  
Qy 3963 TCAAGGCTCAACCAACAGAAATTCGACCTTTCAATTCAGATGAATACCCGTCTGTCG 4022  
Db |||||  
Qy 3979 TTAAGCGCAACCGAAGCAGAACTGGATCTGTCAATTCAGACAGAAATTCAGAGGTGTC 4038  
Db |||||  
Qy 4023 AAACAATTTGTCTACCAATTCGAAGCAGATCAACGGTATTTTGG---CCGGTTTCTCAGAGC 4079  
Db |||||  
Qy 4039 AAACGATTTGTATCATTTCAAGAAAAATCAACGCAATATTGGTCTCTTTTTCAGTGAGC 4098  
Db |||||  
Qy 4080 TTAACAGGTTGTGCTCGAGGCATTTGATTTCTAAGAAAGTTTCTTTTCTTTTACTAGGAAAA 4139  
Db |||||  
Qy 4099 TTAACAGGCAATTTACTTGAAGTATTTGACTCAAGCAGATTTCTTTGTTCTTTTACAGAGAA 4158  
Db |||||  
Qy 4140 CTCGCAACAGATTTCAAGAAATTTTCTCGGATCTCGACTCTCGACGTTCTTATCGATGTGT 4199  
Db |||||  
Qy 4159 CACCGCTCAGATCGAAGATTTCTACGGAGATCTAGACATCATGTGCCCAATGGAACGTAC 4218  
Db |||||  
Qy 4200 TAGAACTGGATTTCTTAAAGTATGATTAAGTCAAGAACGAGTTTCAATGTGTGTAGAGT 4259  
Db |||||  
Qy 4219 TTGAGTTGATGTTTTCGAAGTATGATAAGTCTCAAAACGAGTTTTCATTTGTGCTGTTGAGT 4278  
Db |||||  
Qy 4260 ATCAAAATATGAAAAAGATTTGGGTCTCAATGAGTTTTTGGCCGAGTGTGGAAACAGAGGC 4319  
Db |||||



Db 4279 ACCGAACTCGAGAGACTGGGCTCGAGGATTTCTTGGCAGAAAGTGTGGAAACAAGGGC 4338  
Qy 4320 ACAGGAAACCACTTTGAAGGATTACATTCGTCTGGAATCAAGACATGCTGTGTGATCAAA 4379  
Db 4339 ATAGAAAAACCACTCTGAAAGATTACATCTGCTGTATAAAAACGTGTATTGTATCCAGA 4398  
Qy 4380 GGAAGAGCGGTGATGATGACTCTTTCATCGGCAATACTGTTAATAAGACGCTTGTGG 4439  
Db 4399 GAAAGAGTGGTGTATGTACAACTTTTATCGGTAAATACCGTCATCATTTGCTTGTGTAG 4458  
Qy 4440 GTTCAATGTTACCGATGGAAGAGTCAATAAGGCTCTTTTGTGGAGAGGATTCGGTT 4499  
Db 4459 CATCAATGCTCCGATGGAATAATTGATAAAGAGCCTTCTCGGAGATGACAGTTGT 4518  
Qy 4500 TGTATTTTCCAAAGGTTTGGATTTCCCTGACATTCAGTCAATGCTTAATCTCATGTGA 4559  
Db 4519 TGTACTTTCTTAAGGTTGTGAGTATCCCGATATACAAAGCTGCTAATCTAATGTGA 4578  
Qy 4560 ATTTGAGGCCAACTGTATAGAAAGAGGTACGGTTACTTTTGTGGTAGATACATCATAC 4619  
Db 4579 ATTTGAGGCCAACTGTTCAAGAACCAATATGGGTACTTCTCGGGAGGTACGTGATTC 4638  
Qy 4620 ACCATGATAGGAGCAATAGTGTATATGATCCTTTGAAGTTGATCTCCAACTTGGG 4679  
Db 4639 ATCAGATAGAGGTGTCATAGTATATACACCTTTTGAAGCTGATTTGAAACTTGGTG 4698  
Qy 4680 CAAAAACATATCAAGGATTTATGATCACTTAGAAGATTAAAGGTGTCTTTGTGCGATGTTG 4739  
Db 4699 CTAAACACATCAAGGATTTGGATCACTTTGAGAGGTTTCAAGATCCCTCTGTGTGTTG 4758  
Qy 4740 CTTGTTGCTCGAAACTGTTGCTTAGGCTTTCCGAGCTGAAAGCAGCTATCAAGAGG 4799  
Db 4759 CTGAGTCTGTT--GAACAATTCGCGGTATATACACAATTTGGACGACGCTGTGGGAGG 4815  
Qy 4800 TTCAATAAACCGGATGTAGTGTCTGTTGCTTTTAAATTTGTTTAAACAATTTTCTGTG 4859  
Db 4816 TTCAATAAACCGGCCCACTGTTGCTTTGTTTATAGAGTTAGTTAAGTATTTGTGAC 4875  
Qy 4860 ATAAATTTTATTTAGAACTTTGTTTAAATGGCTTTAGTCTCAGAGATATCTGCAAA 4919  
Db 4876 ATAAAGTTTGTATTAGAGTTTATTTCTGATGCTCTAGTCTTAAAGGTAGGTAAT 4935  
Qy 4920 ATTAGGAGTTTATGATCTTTTGAACAGGATGAGATCTTCCGCAATTCATGACTAAG 4979  
Db 4936 ATTAATGAGTTTATCGATCTGTCAAGTCTGAGAACTTCTCCGCTCGATGTTCAAGCCT 4995  
Qy 4980 GTCAAGGTTTATAGATATCGACTGTGGCAAGATTATGCTGTTAAGATGATGATCTT 5039  
Db 4996 GTAAAGAGTGTATGTTTCAAGGTTGATAGATTATGCTCCATGAAATGAATCATG 5055  
Qy 5040 TCTGATGTAGATTTACTTAAAGGTTTAAAGTTAGTTAAGAAAGGTTATGTTGCTTAGCT 5099  
Db 5056 TCTGAAGTAATCTTTAAAGGTTTAAACTTATAGAGGTTGGTATGTTTCTTAGTC 5115  
Qy 5100 GATTTGTTAGTGTCTGGGAGTGGAACTCTCCGATAACTGCCGTGGTGGTGTGAGTT 5159  
Db 5116 GGTCTTGTGTGTCGGGTGAGTGGAAATTTACCAGATAATTTGCCGTGGTGGTGTGAGTGC 5175  
Qy 5160 TGTATTTAGATTAAGATAAGAAAGAGTAAAGAGTAAAGCAACGCTGGTGGGTATCAGCC 5219  
Db 5176 TGCATGGTTGACAAGAAATGGAAGAGCGGACGGAAGCCACACTCGGGGTATATTAACAT 5235  
Qy 5220 CTTGCTTGCAAAAAAGAAATTTCTTTTAAAGCTAATCCCTAATTTTCAATAACATCCGAG 5279  
Db 5236 GCTGCTGCTTAAAGCGGTTTCAAGTTTAAAGTGGTCCCAATTTACGGTATTAACAAG 5295  
Qy 5280 GATGCTGAGAGCAACCGTGGCAGTGTGTTAGTAATCAAGAGAGTGGCTATGGAAGAA 5339  
Db 5296 GATGCAAGAAAGAACATATGCGAGGCTTATGATAATTAATAATGTAATAAATGAGTGCG 5355  
Qy 5340 GGATACGTCTTTATCTTTGGAGTTCGTTTCAATTTGTTGATGATACATAAAAAATAATGTA 5399

RESULT 11  
AF332868

Db 5356 GGCTACTGCCCTTTGTTCATTAGAAATTTGTGTGTGTGTATTGTTTATAAATAATATA 5415  
Qy 5400 AGAAAGGTTTGAAGGAAACGTATTTTGTAGTGTGACAGCGGCTCCCAATTTGAACCTCACT 5459  
Db 5416 AAATTTGGGTTTGAAGGAAAGTAAACGAGTGTGAACGATGGAGAGCCCATGGAATTCG 5475  
Qy 5460 GAAAGGTTTTCAGGAGTTCGTGGATGAAGTACCAATCGCTGTGTAACCTCGAAAGGTT 5519  
Db 5476 GAAGAAGTTTGTGATGAGTTCATGGAGAAATGTTTCCAATGTCCGTTAGACTCGCAAGTTT 5535  
Qy 5520 CCGGAAAAACAAGAAATGGTAGTAATAATGTTTAAATAAT----- 5561  
Db 5536 CGAACCAATCTCAAAAGAGGTCGGAATAATAATAATAATTTAGTAGGGGCGTTCA 5595  
Qy 5562 -----AGAAATAATAAATACAGTGTGAAGAGGTTTTTAAATTTAGGAAATTTAGGAT 5615  
Db 5596 GCGGAGAGCCATAACCAAAAAAGTTTGTAGTAAAGTTTAAAAAGAGTTTAAATAATTTGAT 5655  
Qy 5616 AATGTAAGTGTAGCAGGCTATCGGTCATCGAGTACGTTTAAATCAATATCGCTTATA 5675  
Db 5656 AAAGATTAAAGCCGAGAGCGTGGTCCGGAATCTGATTCGTTATTA---AATATGCTTACT 5712  
Qy 5676 CAATCAACTCTCCGAGCCAAATTTGTTTACTTATCTTCCGCTTACGAGATCCCTGTGAGC 5735  
Db 5713 CAATCACTTCTCCATCGAATTTGTTTGTCTATCTGTATGGGCTGACCTATAGAAT 5772  
Qy 5736 TGATCAATCTGTGTACAAATGCAATTTGGGTAACCAAGTTTCAACGCAACAGCTAGAGCAA 5795  
Db 5773 TGTAAACGTTTGTACAAATTCGTTAGGTAAACCAAGTTTCAACACAGCAAGCAAGACTA 5832  
Qy 5796 CAGTCAACAGCAATTTGGGATGCTGGAACCTGTGCTAGTATGACAGTACAGATTTTC 5855  
Db 5833 CTGTTCAACAGCAGGTTTCAAGAGGTTGGAAACCTTTCCCTCAGAGCACCGTCAGATTTTC 5892  
Qy 5856 CTGATCGGATTTCTATGTTATAGATATAATTCGACGCTTGTATCGTTGATCATCGGCGT 5915  
Db 5893 CTGGCGATGTTTATAAGGTTACAGGTACATGACGATTTTAGATCTCTTAATTTACTGCGT 5952  
Qy 5916 TATTAAATAGCTTTGTATCTAGAAATAGAAATATAGAGGTTGATAATCAACCCGCAACCGA 5975  
Db 5953 TGCTGGGGCTTTTGATACTAGGAATAGAAATATCGAAGTAGAATAACACAGCAGAGTCCGA 6012  
Qy 5976 ATACTACTGAATTCGTTAAAGGACTCAGAGGTTAGACGATGCTACTGTAGCTATAAGG 6035  
Db 6013 CAACAGCTGAAACGTTTAGATGCTTACCCGAGGGTAGACGCTACGGTTGCAATTCGGT 6072  
Qy 6036 CTTCAATCAATATTTTGGCTTAATGAACTGTTCTGCGAACTGGCATGTTTCAATCAAGCAG 6095  
Db 6073 CTGCTATAAATTAATTTAGTTTAACTAGTAAGAGTACTGACCTGTACATCAGNATA 6132  
Qy 6096 GCTTTGAGACTGCTAGTGGACTTGTCTGGACCAAACTCCGGCTACTTACTAGCTATTGTTGT 6155  
Db 6133 CTTTGAAGTATGCTGGGTTGGTCTGGACCTCTGCACCTGCATCTTAAATCATAGGT 6192  
Qy 6156 GAGATTTCCCTAAATAAAGTCGCTGAAGACTTTAAATTCAGGTTGCTGATACCAAAATC 6215  
Db 6193 GCTGAAATATAA-----GTTTGTGTTTCTTAAACACACACGTTGTCGTACGATAACG 6244  
Qy 6216 AGCAGTGGTGTGTCGCTCACTTAAATATAACGATTCATATCTGGATCCAACTGTTAAA 6275  
Db 6245 TACAGTGTGTTTCCCTCCACTTAAATTCGAAGGGTAGTGTCTTGAGGCGCGGAGTAAAC 6304  
Qy 6276 CCATGTGATGGTGTATATCTGTGTATGGGTGTAACATCGGAGAGGTTTGAATCTCTCCC 6335  
Db 6305 ATATATGGTTTCATATATGTCGCTAGGCACGTAAAAAAAGCGAGGATTCGAATTCCTCCG 6364  
Qy 6336 TAACCCCGGTAGCGGCCA 6355  
Db 6365 GAACCCCGGTGGGGCCCA 6384

LOCUS AF332868 6383 bp ss-RNA linear VRL 12-MAR-2000  
 DEFINITION Tomato mosaic virus isolate Queensland, complete genome.  
 ACCESSION AF332868  
 VERSION AF332868.1 GI:13274576  
 KEYWORDS  
 SOURCE Tomato mosaic virus  
 ORGANISM  
 Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.  
 REFERENCE 1 (bases 1 to 6383)  
 AUTHORS Lee, S.M., Smith, G.A. and Polkinghorne, I.G.  
 TITLE Submitted (29-DEC-2000) Department of Biochemistry, The University  
 of Queensland, QABC, Gehrmann Labs, The University of Queensland,  
 St. Lucia, Brisbane, QLD 4072, Australia  
 JOURNAL  
 COMPLETE Complete nucleotide sequence of tomato mosaic virus Australia  
 isolate  
 UNPUBLISHED  
 2 (bases 1 to 6383)  
 AUTHORS Lee, S.M., Smith, G.A. and Polkinghorne, I.G.  
 TITLE Direct Submission  
 JOURNAL  
 FEATURES  
 source  
 1..6383 /organism="Tomato mosaic virus"  
 /mol\_type="genomic RNA"  
 /isolate="Queensland"  
 /specific\_host="Lilicopersicon esculentum cv. Grosse Lisse"  
 /db\_xref="taxon:12253"  
 /country="Australia: Queensland"  
 72..4922 /note="involved in genomic replication; contains  
 readthrough stop codon"  
 /codon\_start=1  
 /transl\_except=[pos:3420..3422,aa:OTHER]  
 /product="184 kDa replicase"  
 /protein\_id="AAK17989.1"  
 /db\_xref="GI:13274578"  
 /translation="MAYTQTATSSALLETVRGNTLVNDLAKRLRYDVTADFEFNARDP  
 RPKVNFSGVSEETQITATKAYPEFOITFNTQNAVHSLAGLSAGLSLELYLMMQVHP  
 SLVTDIGENFASHLPKGRVAVCHCMPLNDVRDIRHVGOKDSTELVLSRLERGNKHVP  
 NFQKEAFDRVAEPNEVYVCHDTQTCRHSOECYTGVRVATLHSHIVDIPADEFGAALL  
 RKNVVCYAPAFHSENNLLSDSHVLDENINACQORDGLRTTFSESTLNYHSHSVN  
 ILKVCTYTPFASNREYMERKEFLVTRVNTWCFPSRIDTFLLYKGVAHGKVDSEQFTK  
 AMEDAHYKKTLCAMNSERISLEEDSSVNTWFFMRDMVILPFLDISLSTSKRTRKEV  
 LVSKDFYTVLNIHRTYQKALTYSNLVSESRVINGVTAARSEMDVSKLLQS  
 LSWTFPLTKLVLKDDLLSKALPGKTSQVHWDEISLAFGNAPTSIKERLINRKL  
 IKITENALETRVPLDYTHDRVLSEYKMSVMPVLDIRKKMEETEMYNALSELVIL  
 KNSDGPVDFVSQMCQSLSEYDPMPTAAKVIYVMSNESGLTLTPEQPTANVALAODS  
 EKASDGLVDFVSQMCQSLSEYDPMPTAAKVIYVMSNESGLTLTPEQPTANVALAODS  
 TASGLIHQKQSVIYVTPGLVQKQNFIDLSVASLSAAVNSLVKILKDTALDLETRQ  
 KFGVLVDYSAKRWLVPKSAKNHGWETHARKYHVALLEHDEFGIITCDNRRVAVSS  
 ESVVYSDMAKLRTLLRLKQDEHPHVSAAKVLVDVPGCGKTEILSRVNFEBDLILV  
 PGQAASMIARRANASGIIIVATQDNVTVDFLMNYGKGARCOFKRLFIDEGMLMHTG  
 CVNPLFWSLSDIAYVGDYQOILPYINRVTGPPYPAHFAXLEVDETRRTTLRCAPD  
 VTHFLNORYEGHVWCTSESBKSKSQEMVSGAASINPVSKPLKGILTFQSDKEALLS  
 RGVADVHTVETHPQDVSILVKRLTPVSIITARDSPHVLVSLSRSHSLKTYTVVM  
 DPLVSIITRDLSEVSSYLDDMYKVDAGTQXQDQVDSVFNKNFLVFAAPKTGDI5DMQY  
 YDKLPGNSLTLLNNDYATWKLITDILSNVXDCIILDKMSVAAKPDVKPTLIPMVRTAA  
 EMPROTGLLENLIVAMIKNFNSPELSDVPIENTASLVVDKFPDYSLLLEKRPKNKFN  
 SLFRESLRSNWIAKQEOVTIGOLADPDFVDLPADQVVRHMIKAQPKOKLDLSIQTEVP  
 ALQTIYVHKKINAI FGPLSELTRQLLSDSDSRFLFPFTRKTPAQI0EDFPGDLDSHV  
 PMVLELDGSKYDQKQNEFTCAVEIETWRRLGLEDLEAVNKGQHRKTTIKDYTAGIK  
 TCDWLEQKSGDVTFFIGNTVILASCLASMLPMEKLIKGAFCGDDSLLYFPKGCEYDPI  
 QORANLWNFEAKLPKQYGFYGRVYVTHDRGCIYVDYPLKLSKLGAXHDKDWDLH  
 EEFRSRLCDVAESLNNCAVYITQLDDVAGVEVHKTAPPGSFVYKSLVKYLSKDLFRSLF  
 LDGSCC"  
 72..3422  
 /note="involved in genomic replication"  
 /codon\_start=1  
 /product="126 kDa replicase"  
 /protein\_id="AAK17989.1"  
 /db\_xref="GI:13274577"  
 /translation="MAYTQTATSSALLETVRGNTLVNDLAKRLRYDVTADFEFNARDP  
 RPKVNFSGVSEETQITATKAYPEFOITFNTQNAVHSLAGLSAGLSLELYLMMQVHP  
 SLVTDIGENFASHLPKGRVAVCHCMPLNDVRDIRHVGOKDSTELVLSRLERGNKHVP  
 NFQKEAFDRVAEPNEVYVCHDTQTCRHSOECYTGVRVATLHSHIVDIPADEFGAALL  
 RKNVVCYAPAFHSENNLLSDSHVLDENINACQORDGLRTTFSESTLNYHSHSVN  
 ILKVCTYTPFASNREYMERKEFLVTRVNTWCFPSRIDTFLLYKGVAHGKVDSEQFTK  
 AMEDAHYKKTLCAMNSERISLEEDSSVNTWFFMRDMVILPFLDISLSTSKRTRKEV  
 LVSKDFYTVLNIHRTYQKALTYSNLVSESRVINGVTAARSEMDVSKLLQS  
 LSWTFPLTKLVLKDDLLSKALPGKTSQVHWDEISLAFGNAPTSIKERLINRKL  
 IKITENALETRVPLDYTHDRVLSEYKMSVMPVLDIRKKMEETEMYNALSELVIL  
 KNSDGPVDFVSQMCQSLSEYDPMPTAAKVIYVMSNESGLTLTPEQPTANVALAODS  
 EKASDGLVDFVSQMCQSLSEYDPMPTAAKVIYVMSNESGLTLTPEQPTANVALAODS  
 TASGLIHQKQSVIYVTPGLVQKQNFIDLSVASLSAAVNSLVKILKDTALDLETRQ  
 KFGVLVDYSAKRWLVPKSAKNHGWETHARKYHVALLEHDEFGIITCDNRRVAVSS  
 ESVVYSDMAKLRTLLRLKQDEHPHVSAAKVLVDVPGCGKTEILSRVNFEBDLILV  
 PGQAASMIARRANASGIIIVATQDNVTVDFLMNYGKGARCOFKRLFIDEGMLMHTG  
 CVNPLFWSLSDIAYVGDYQOILPYINRVTGPPYPAHFAXLEVDETRRTTLRCAPD  
 VTHFLNORYEGHVWCTSESBKSKSQEMVSGAASINPVSKPLKGILTFQSDKEALLS  
 RGVADVHTVETHPQDVSILVKRLTPVSIITARDSPHVLVSLSRSHSLKTYTVVM  
 DPLVSIITRDLSEVSSYLDDMYKVDAGTQXQDQVDSVFNKNFLVFAAPKTGDI5DMQY  
 YDKLPGNSLTLLNNDYATWKLITDILSNVXDCIILDKMSVAAKPDVKPTLIPMVRTAA  
 EMPROTGLLENLIVAMIKNFNSPELSDVPIENTASLVVDKFPDYSLLLEKRPKNKFN  
 SLFRESLRSNWIAKQEOVTIGOLADPDFVDLPADQVVRHMIKAQPKOKLDLSIQTEVP  
 ALQTIYVHKKINAI FGPLSELTRQLLSDSDSRFLFPFTRKTPAQI0EDFPGDLDSHV  
 PMVLELDGSKYDQKQNEFTCAVEIETWRRLGLEDLEAVNKGQHRKTTIKDYTAGIK  
 TCDWLEQKSGDVTFFIGNTVILASCLASMLPMEKLIKGAFCGDDSLLYFPKGCEYDPI  
 QORANLWNFEAKLPKQYGFYGRVYVTHDRGCIYVDYPLKLSKLGAXHDKDWDLH  
 EEFRSRLCDVAESLNNCAVYITQLDDVAGVEVHKTAPPGSFVYKSLVKYLSKDLFRSLF  
 LDGSCC"  
 72..3422  
 /note="involved in genomic replication"  
 /codon\_start=1  
 /product="126 kDa replicase"  
 /protein\_id="AAK17989.1"  
 /db\_xref="GI:13274577"  
 /translation="MAYTQTATSSALLETVRGNTLVNDLAKRLRYDVTADFEFNARDP  
 RPKVNFSGVSEETQITATKAYPEFOITFNTQNAVHSLAGLSAGLSLELYLMMQVHP  
 SLVTDIGENFASHLPKGRVAVCHCMPLNDVRDIRHVGOKDSTELVLSRLERGNKHVP  
 NFQKEAFDRVAEPNEVYVCHDTQTCRHSOECYTGVRVATLHSHIVDIPADEFGAALL  
 RKNVVCYAPAFHSENNLLSDSHVLDENINACQORDGLRTTFSESTLNYHSHSVN  
 ILKVCTYTPFASNREYMERKEFLVTRVNTWCFPSRIDTFLLYKGVAHGKVDSEQFTK  
 AMEDAHYKKTLCAMNSERISLEEDSSVNTWFFMRDMVILPFLDISLSTSKRTRKEV  
 LVSKDFYTVLNIHRTYQKALTYSNLVSESRVINGVTAARSEMDVSKLLQS  
 LSWTFPLTKLVLKDDLLSKALPGKTSQVHWDEISLAFGNAPTSIKERLINRKL  
 IKITENALETRVPLDYTHDRVLSEYKMSVMPVLDIRKKMEETEMYNALSELVIL  
 KNSDGPVDFVSQMCQSLSEYDPMPTAAKVIYVMSNESGLTLTPEQPTANVALAODS  
 EKASDGLVDFVSQMCQSLSEYDPMPTAAKVIYVMSNESGLTLTPEQPTANVALAODS  
 TASGLIHQKQSVIYVTPGLVQKQNFIDLSVASLSAAVNSLVKILKDTALDLETRQ  
 KFGVLVDYSAKRWLVPKSAKNHGWETHARKYHVALLEHDEFGIITCDNRRVAVSS  
 ESVVYSDMAKLRTLLRLKQDEHPHVSAAKVLVDVPGCGKTEILSRVNFEBDLILV  
 PGQAASMIARRANASGIIIVATQDNVTVDFLMNYGKGARCOFKRLFIDEGMLMHTG  
 CVNPLFWSLSDIAYVGDYQOILPYINRVTGPPYPAHFAXLEVDETRRTTLRCAPD  
 VTHFLNORYEGHVWCTSESBKSKSQEMVSGAASINPVSKPLKGILTFQSDKEALLS  
 RGVADVHTVETHPQDVSILVKRLTPVSIITARDSPHVLVSLSRSHSLKTYTVVM  
 DPLVSIITRDLSEVSSYLDDMYKVDAGTQXQDQVDSVFNKNFLVFAAPKTGDI5DMQY  
 YDKLPGNSLTLLNNDYATWKLITDILSNVXDCIILDKMSVAAKPDVKPTLIPMVRTAA  
 EMPROTGLLENLIVAMIKNFNSPELSDVPIENTASLVVDKFPDYSLLLEKRPKNKFN  
 SLFRESLRSNWIAKQEOVTIGOLADPDFVDLPADQVVRHMIKAQPKOKLDLSIQTEVP  
 ALQTIYVHKKINAI FGPLSELTRQLLSDSDSRFLFPFTRKTPAQI0EDFPGDLDSHV  
 PMVLELDGSKYDQKQNEFTCAVEIETWRRLGLEDLEAVNKGQHRKTTIKDYTAGIK  
 TCDWLEQKSGDVTFFIGNTVILASCLASMLPMEKLIKGAFCGDDSLLYFPKGCEYDPI  
 QORANLWNFEAKLPKQYGFYGRVYVTHDRGCIYVDYPLKLSKLGAXHDKDWDLH  
 EEFRSRLCDVAESLNNCAVYITQLDDVAGVEVHKTAPPGSFVYKSLVKYLSKDLFRSLF  
 LDGSCC"

S L T Y D I G G N F A S H L F K R A Y V H C M P N L D V R D I M B H E G Q K D S I E L Y L S R L E R G N K H V I  
N F Q K A F A R A E M P N E V C H D T F Q I C R S H Q E C T G R V A Y A I A H S I Y D I D I P A D E F G A A L  
R K N V H C Y A A F S N L L E D S H N L D E I N A C F O R D D R L T F S F A S E S I N T S H S Q Y K  
I L K Y C K V T F P A S N E V M K E L V T R W F C K F S R I D T F L L Y K G V A H G K V D S E Q Y K  
A M E D A H Y K K T L A M C N S B R I L L E D S S V N Y W F P K M R D M V I V P L D I S L T S K R T R X E V  
L V S K P F V T V L N H I R T Y Q A K A L T Y S N V L S F V E S I R S R V L I N G V T A R S E W D V K S L Q S  
L S M T F L H T K L A V L K D D L L S I F A L P G K P K T V S Q W V D B I S I A F A G N A P T S K E R L I N R K L  
I K T I N A L E I R V P D I Y V T F H D L S E Y K M S V M P V L D I R K M E E T E E M N A J S E L S V L  
K N S D F V D T S Q M C Q S L E D P M T A K V T V A M S N E S G I L T F E Q T E A N V A L A Q S  
B K A S D G A L V T S R D V E E P S I K G S M A R G E L Q A L G S D G V P E S S Y T R G E E I E S L E Q F H M A  
T A S S I L H K M C S I V T G P L K V O O M K N F I D S L V A S I S A A V S N L V K I I L K D T A A I D L E T R Q  
K F G L D V A S K R M L V P R S A K N H A N G V E T H A R K H V A L L E H D E F G I T C D N W R V A S V  
E S V Y D S M A K L T R L R K L D G G B H V S S A K V I V D G V P G G K T K E I I U R V N F E S D L I V  
P G R A E M I R R A N S A G I I V A T K D N V T V D S F L M N Y K G A R C Q F K L F I D E G L M L H T  
Q C F V E M S L C D I A Y V Y G T Q O I P Y I N R V T G P Y P A H F A K L E V D E V E T R T L T C D P A D  
V T H F L N O R V E G H W M C T S S B K S V S O E M S V G A A S I N P V S K P L K G I I L T F T Q S D K E A L L S  
R G Y A D V H T V E H Q E T A D S V L V R L T P T P V S I I A R D S P H V I V S L S R H T K S L K Y T V T V M  
D P E V S I I R D L E R S S Y L L D M V L D A G T Q

CDS 4906..5700

/note="involved in cell-to-cell movement of virus in plants"

/codon\_start=1

/product="30 kDa movement protein"

/protein\_id="AAK17990.1"

/db\_xref="GI:13274579"

/translation="NALVVVGKVNIINEFDLSKEKLPSMFTPTPVMSVMYVKVDKMIVHESLSVNLLKGVLIGBVYCLVGLVVSGEWNLPDNCRGVSVCMVDKRMRADERATGCVYTAAAKRFQFKVPNGVIQTDAEKINWOVLNIKNVMSAGYCPLSLEFVATCISYVTKNIIKLGRKYTSNDGPMELSSEVDFEMENPMVSLAKEPRITKSXRGSYCNVNLGNKGRSGRSKPSPKFSFEVEKEFNLDIEDEAETSVADSDSY"

CDS 5703..6182

/note="involved in virus assembly and systemic spread of virus in plants"

/codon\_start=1

/product="17.5 kDa coat protein"

/protein\_id="AAK17991.1"

/db\_xref="GI:13274580"

/translation="MSYSITSQSQFVFLSSWADPIELMLNVCTNSIGNOFQTOAQRTVOQOSWEKVPPOSTPPGVYRVYNALVDPLITALLGAFDTRNR.IIEVENOQSPTTAETLDATRVD DATVAIRSAINNLVNLVRGTGLYNQNTFESMSGLVWTSAAPS"

ORIGIN

	Query Match	Best Local Similarity	Mismatches	Conservative	Score	DB 14;	Length	6383;
Qy	18	CGACACAACAATTAAAACAAAAACAACATATTATACAAACAACAACAACAACAATGGCAC	77					
Db	23	CAACAACAACAACAACAACAATTTACATTTTACATTC---TACAACTACAAATGGCAT	78					
Qy	78	ACATACAACTATAATTAGCAACGCCCTTCTTGAAAGCGTGAGTGGTAAAAACAACCTCTCG	137					
Db	79	ACACAACAACAGCCACATCGTCCGCTTTTGCTTGAGACCGTCCGAGGTAAACAATCTTGG	138					
Qy	138	TTAATGACCTTGCAGAAGCGCATGTACGATACGGCCGTGGGAAGAAATTTAACGCCCGCG	197					
Db	139	TCAAAGATCTTGACAGCGCGTCTATATGACACACGCGTAGATGAATTTAATGCTAGG	198					
Qy	198	ACCGTAGACAAAGGTCACATTTTTCCAAACTATTAGCGAAGAGCAACCGCTTCTAGTCT	257					
Db	199	ACCGCAGGCCCTCAAGTCAATTTTTTCCAAAGTAGTAGCGAAGAACACAGCGCTATTTCGAA	258					
Qy	258	CGAACGGTACC GGAGTTCACATTTTATATATCTAAAAATGCGGTACACAGTT	317					
Db	259	CCAAGCCTACCCAGAAATTCCAAATTAATCTTCAACAACGAGAAATGCTGTGATTTCCC	318					
Qy	318	TGGCTGGAGGTTTGAGAGCATTAAGAATTGAAATATCTGATCTCAAGTTTCCCTATGGAT	377					
Db	319	TTGAGGCGGTCTCCGATCATTAGATTGGAATATCTGATGATGATCAATAATCCCTACGGAT	378					
Qy	378	CGCCGACATATGATATAGTGGGAACTTTTCGACGACATTTTGTTCAAAGCGAGGATACG	437					
Db	379	CATTGACATATGATATCGAGGTAAATTTTCGATCTCATCTGTTTCAAAGGCGGACATACG	438					

QY 438 TGCATTGCTGTATGCCCAATCTCGACATACGAGATATAATGAGGACGAAAGGCAAAGG 497  
DB 439 TTCACTGCTGTATGCCCAATCTAGATGTCGCGACATAAATGCGGACGAGGCGCAAAGG 498  
QY 498 ACTCAATTTGAGATGTATTTCTGTCAGATTTGTCGTTCTTAACAAGGTAAATTCCTGAGTTTC 557  
DB 499 ACAGTATTGAACATATACCTTTCTAGGCTCGAGAGGGCAACAACATGTCCTCCAAACTTCC 558  
QY 558 AAAGGGAGGCTTTTAAACAGATATGACAGAGCTCCCAAGAGTCTGCTGCTCTTAAACTT 617  
DB 559 AAAGGAAGCTTTTCGACAGATACGCTGAATGCGCAAGAGTAGTCTGTACAGTACTT 618  
QY 618 TTCAAGATTTGATATATATTCCTGTCAGATGTCGTCGTTTGAAGAGTATGCTGCTGCTG 677  
DB 619 TCCAAACGCTGAGGCATTTCTCAAGATGTTTACCGGAAGAGTGTATGCTATTGCTTTGC 678  
QY 678 ACAGTTTGTATGATATATTCCTGTCAGATGTCGTCGTCGTCGTTTGAAGATATATC 737  
DB 679 ATAGTATATACGATATACCTTGCAGAGAGTTCGCGCGGCACCTGCTGAGAAAGATGTAC 738  
QY 738 ATGTATGTTATGACGCTTCCATTTTGGCAGAAGCATTTATTAAGACACGAGGTTA 797  
DB 739 ATGTATGTTATGCGCTTTCCACTTTTCCGAGATTTACTTCTCGAAGATTCACAGTCA 798  
QY 798 CGCTTAATGAATAGGCGCACTTTTCAAAGAGAGGTGATGATTTTCTTTTCTTTG 857  
DB 799 ACCTCGATGAGTCAATGTCATGTTTCCAAAGAGATGAGACAGGTGACCTTTTCTTTG 858  
QY 858 CTGATGAAGTACTTTTAAATTTATGATCATTAATACAAATATATCTTGCATTTAGTTA 917  
DB 859 CATCTGAGTACTCTTAAATTTATGATCATAGTTATTTCTTAATATTTCTTAAGTATTTGCA 918  
QY 918 AATCTTACTTTCTGCTTCTAGTAGAATGTTTACTTTAAGGAATTTTATGCTCACTAGG 977  
DB 919 AACTTACTTTCCAGCTCTTAATAGAGGTTTACATGAAGGATTTTATGTAAGTATAG 978  
QY 978 TTAATCTTGGTTTGTAAATTTTCAAAGTAGATACCTATATTTCTGTACAAGAGTTA 1037  
DB 979 TTAATCTTGGTTTGTAAATTTTCTAGAAATAGATATCTTTCTTATTGTACAAGAGTTAG 1038  
QY 1038 GACAAGTGGGTGTAGTATGATGATCTTATGAGCGATGGAAGCGCTTTGCTTACA 1097  
DB 1039 CGCATTAAGGTTGATAGTATGAGCGTTTACAAGGCTTATGGAAGCGCATGCGACTACA 1098  
QY 1098 AGAAACCTTTGGCATGTTTCAAACACTGAAGAGCAATCTTTAGAGACACGGCTTCGGTTA 1157  
DB 1099 AAAGACTCTTGGATGTGCAACAGTGAAGAAATCTTTGTAGAGATTTCTTATCAGTTA 1158  
QY 1158 ACTTTTGGTTTCCCTTAAGATGAAGACATGTTGATGATPACCGCTGTTTGGGGTTCTATTA 1217  
DB 1159 ATTACTGGTTTCCAAATGAGGATATGGTATAGTTTCCACTATTTTGACATATCTCTCG 1218  
QY 1218 CCAGCAAAAGATGACAAGAGTGAAGTCAATGTTTAACTGATCTGTTTACACAGTGC 1277  
DB 1219 AGACTAGTAAAGAACACGCAAGAGGCTTTAGTTTCAAAGGCTTTGTTTATACAGTGT 1278  
QY 1278 TTAATCATATCAGAAATATCAAGCCAAAGCGTTTAACTTACCAGAACGTATTTATCTTTCG 1337  
DB 1279 TAAATCATTTGATCGTACAGCCAAAGCGCTTACTTACTCCAGCTGTTATCTTTTCG 1338  
QY 1338 TGGAGTCTATAAGATCCCGGTGATATCAATGGTGTACTGCTAGGTCTGTAATGGGATG 1397  
DB 1339 TCGAATCAATTCGTTTCGAGAGTGATCATTAACGGGTTACTGCCAGGTCTGAGTGGGATG 1398  
QY 1398 TAGATAAGCAATTTCTTCAACCTTTGTAAGTACTTTCTTCTGACAGACTAAGCTGGCTG 1457  
DB 1399 TCGATAAATCATTTATACAGTCTTGTGATGAGCTTTCTTCTTACATACCAAGCTTGGCG 1458  
QY 1458 CGCTTCAAGACGATATAGTAAATGGAAAGTTTGGTCTTGGATTAAGACCACTTCTGAAC 1517  
DB 1459 TTCTGAAGACGATCTTTTGTATGCAAGTTTGCATTTGGATCTTGGACCAAAACTGTCTCAAC 1518

QY 1518 TTATTTGGGATGAGGTGGGCAAAATTTTGGAAACGTTTCCCACTATCATCAAGAGAGAT 1577  
DB 1519 ATGTGTGGGATGAGATTTCCCTAGCTTTCGGCAATGCTTTCCCATCGATCAAGAAAGAT 1578  
QY 1578 TGGTGACGAGGAAATTTCTGGATGTAAGTGAGATGCTCTGAAGATCAAGATCCAGATC 1637  
DB 1579 TGATAAACCGGAACTGATCAAAATTTACGGAGATTTGGTTAGAGATCAGGGTCCCGATC 1638  
QY 1638 TGTATGTCAATGGAAGACAGGTTTCGTAGCTGAATACACAAAGTCTGAGAGTTACCGC 1697  
DB 1639 TTTATGTCACTTTCCATGATAGGTAGTTTCTGAGTACAAATGTCAGTGCAGATGCGG 1698  
QY 1698 ATCTAGATATCAAGAGGACTTAGAAGAGCTGAGCAATGTTAGGACGCTTATCAGAA 1757  
DB 1699 TGTAGACATTTAGGAAAAGATGGAAGAACTGAGGAAATGTTCAATGCTGCTGCCAAC 1758  
QY 1758 TATCTATCTTAAGGCTGCTGATAATTTTCGATATCGCAAGTTTCAAGACATGTGCAAG 1817  
DB 1759 TGTCTGTACTTAAAAATTCAGACAGTTTCTGATGTTGATGTTTCCAGATGTGCCAAT 1818  
QY 1818 CTTTATAGTGTAGTCTCTGATGTCGACAGCAGTAAATCGTTGTCAGTGGCCGAGAAAGAA 1877  
DB 1819 CTTTAAAGGTTGATCCAAATGACTGCAAGAAAGGTAATAGTAGCAGTTATGAGCAACGAGA 1878  
QY 1878 GCGGTTTAACTCTTACTTTTGTATAGCCACCGAGGAGATGTCGCTTAAGGCTTTAAA- 1936  
DB 1879 GTGGTCTTACTCTCACGTTTGAACAGCCCAACCGAAGCTAATGTTGCGCTAGCATTCGAAG 1938  
QY 1937 ---- AGCACGGGCTCTGAGGCGTGTATGCTTGAACCGACATCCGAAGAGGTGAACG 1991  
DB 1939 ATTCTGAAAAGGCTTCTGATGGGCGTTGGTAGTTTACCTCAAGAGATGTTGAGGAACCGT 1998  
QY 1992 TAAATAAATTTCTATTGCTGGAAGAGGAGATTTGCCCTGTGTGTGTCGAGAAAGTCAATGTT 2051  
DB 1999 CCATAAAGGTTGATGTCGCGCGGTGAGTTACAAATTTGGCCGATATTCTGGGACGCTC 2058  
QY 2052 TGAGGAATGCTTAATCTTAGAGCACGAGGTTGAGTCCCTCAACGATTTCCATTAAGGCTT 2111  
DB 2059 CTGAATCTTTCATACACTAGGAGCGGAGATGAGTCTCTCGAGCAGTTTCATATGGCAA 2118  
QY 2112 GCGTGATAGTGTATTACAAAGCAATGSCATCGGTTGTCTACACTGGCTCACTCAAG 2171  
DB 2119 CAGTATGCTTAAATTCATAGCAGATGTTTCGATCGTGTACACGGGCCCTCTTAAAG 2178  
QY 2172 TTCAACAAATGAAGAACTATGTGACAGTTTGGCAGCTTCGTTGTCCGCCACTGTATCAA 2231  
DB 2179 TTCAACAAATGAAGAACTTTATAGACAGCTGTGTAGCTCGCTCTCTGCTGCGGTGCGA 2238  
QY 2232 ATCTATGCAAGTCACTAAAGGATGAAGTGGGTATGATTTCTGATTCAGGGAGAAAGTTG 2291  
DB 2239 ATCTAGTGAAGATCTTAAAGATACAGCCGCGATTTGACCTTGAACCTCGTCAAAAGTTCG 2298  
QY 2292 GTGTTTGGGATGTCATTTTGAAGAGTGGCTCTCAACCTCGCGGCAAGGTCATTCAT 2351  
DB 2299 GAGTTCTGAGTGTGCTTGGAAAGGTGGCTAGTTAAACCTCCGCAAGAACCATGCAT 2358  
QY 2352 GGGGAGTGTCTGAGTATCAAGGGGAAATGTTTACTGCACTTCTATCTTATGAAGAG 2411  
DB 2359 GGGGGGTTGTGAGACTCATGCGAGGAATATACCGTCGCATTTACTGGACACGATGAAT 2418  
QY 2412 ATGAATGTGATGAGAGCGACTGAGGAGGGTGGCTGTATCATCTGTATCAATGTGTAT 2471  
DB 2419 TTGGCATTTATGCTGCGATTAATGCGACGGGTGGCTGTGAGTTCTGTAGTCCGCTAGTAT 2478  
QY 2472 ATTCTGATATTCAGAGCTCCAAATCTCAGGAAACAAATGAGAGCGGTGAACCCACG 2531  
DB 2479 ATTCTGATATGGCTAAACTCAGGACTCTTGAGAAAGATTGCTCAAGAGTGGAGAACCA 2538  
QY 2532 AACCTACTGCAAGATGCTACTTGTGGATGGGCTGCTGTTGTTGGAAGTGTACAAAGAG 2591  
DB 2539 TTAGTTTCAGCAAGAGGTGTTTGTGATGAGGCTTCCAGGGTGGGGAAGCAAGGAAA 2598  
QY 2592 ATTTGAAAGATTTGATCTTGTATGAGGATTTTGTATCTTGGTTCTTGGAAAAACAAGCTGCTG 2651

Db 2599 TTCTTTGAGAGTAAATTTTGAAGAAGATCTAATTTCTGTCCCTGGTCTGTCAGAGCTGCCG 2658  
Qy 2652 CTATGATCAGAAGAGGGCTAAATTCATCTGGAATGATGAAGAGCCAAATGCAATGTGA 2711  
Db 2659 AGATGATCAGAAGAGAGCTAAATCGTCGGGCATAATAGTGGCTACAAAGGATAATGTGC 2718  
Qy 2712 GAACGGTAGAATTCACCTTCTAATGATCCAAAACCGCGATCAC-----ACAAGAGGC 2762  
Db 2719 GCACCGTCGATTCATTTTGTGATGAATTAACGGGAAGGGCACGCTGTCAGTTCAAAAGAT 2778  
Qy 2763 TTTTATATGATGAAGGGTGTGATGCTGCACACCGGTTGTGTAACTTCTCGTGGTCTTATCT 2822  
Db 2779 TGTTCATAGACGAAGGTTTGTATGTCATATCTGTTGTGTGAATTTCTTGGTTGAAATGT 2838  
Qy 2823 CTGGTTGCGACATCGCATACATTTACGGAGATACACAGCAGATTCCTTTCAATTAACAGAG 2882  
Db 2839 CTCTGTGCGATATGCGATATGTTTATGGAGACCCCAACAAATTCGCTATCATCAACAGAG 2898  
Qy 2883 TTCAGAAATTTCCGGTATCCCAAAACATTTTGAGAGCTGCAAGTGGATGAATTTGAGATGA 2942  
Db 2899 TAACTGTTTCCCGTACCTCGCGACCTTTGCAAAATTTGGAGGTGCGACGAAGTCGAAACAA 2958  
Qy 2943 GGAGGACCACTGAGAGATGCCAGGTGATGTGAATTTTTTCTTACAAATCGAAGTACGAAG 3002  
Db 2959 GAAGACTACTCTTCGCTGTCGGCTGATGTCAACACTTCTCTAAATCAAAGGTATGAAG 3018  
Qy 3003 GAGGGTGACNAACACTTCACTGTACACGATCGGTCTCATCTGAGATGATAGCGCGTA 3062  
Db 3019 GACACGTAATGTGCACGTCTCTGAAAGAAATCAGTTTTCCAGGAAATGTTAGTGGGG 3078  
Qy 3063 AGGAGTACTAAAACAGTGTGTTTCAAACCACTTAAAGGGAATTTGTAACCTTTCACTCAGG 3122  
Db 3079 CTGGCTCTATCAATCTCTGTCGAAGCGCTTAAAGGGAATTTTGACCTTTCACACAGT 3138  
Qy 3123 CTGATAAATTTGAGTTAGAGAGAGGGCTATAGAATGTGAACACCGTTCATGAGATCC 3182  
Db 3139 CTGACAAAGGAGGCCCTTCTCTCAAGGGGCTACGCAGATGTCCATCTGTATCATGAGGTAC 3198  
Qy 3183 AAGGAGAAACCTTTGAAGATGTGTCGCTGCTGAGTTCAGGCAACTCCACTGACTCTGA 3242  
Db 3199 AAGGTGAGACTTATGACAGAGTATCGTTAGTTCGACTTAAACACTACCGCTGTATCTATCA 3258  
Qy 3243 TTTCCAAAGTCTTCCCGCATGTTCTAGTTCGCTCTGACTAGACACACAAAGAGCTTCAAT 3302  
Db 3259 TCGCAAGAGACAGTCCGCATGTTCTGCTCTGTTGTCAGAGACACACAAATCCCTAAAGT 3318  
Qy 3303 ATACACGGTAGTGTAGATCCCTTTAGTACAGATAATTAGTGTGTTCTTCTTTAAGCT 3362  
Db 3319 ACTACACCGTTGTATGGATCTTTTAGTTAGTATCATTTAGAGATTTTGAACGGGTTAGTA 3378  
Qy 3363 CCTTCTCTTTAGAAAATGTATATGTAGAGCAGGTAGTAGATAGCAATTCAGATGGATG 3422  
Db 3379 GTTACTTATTAGACATGTACAAAGTAGTGCAGGTACTCATAGCAATTCAGGTGCACT 3438  
Qy 3423 CAGTGTTCAAAGGTCAATAATCTTTGTGGCAACACCTTAATTCAGGAGACTTTCCAGATC 3482  
Db 3439 CTGTGTTTAAAAATTTTCAATCTTTTGTAGCAGCTCCAAAGACTGGAGATATATCTGATA 3498  
Qy 3483 TACAGTTCTATACAGATGATGCCCTCCCTGGTATAGTACTATACTTAAACAGATGATG 3542  
Db 3499 TGCAATTTTACTATGATAAGTGTCTCTCTGGGAACAGCACGCTGTTTGAACAACTACGACG 3558  
Qy 3543 CTGTTTACCATCAGGTTTACGTGATAATAGTCTTAAATGGAAGGATGTGTTCTTCAATTTT 3602  
Db 3559 CTGTTTACCAAGAAATGTACTGACATTTCTCTGAATGTCAAGATGTGCATATTAGATAGT 3618  
Qy 3603 CCAAAAGTATTCGATGCAAGAGGAGTGAACCATGTCTAGAGCCAGGTTTTCGTACCG 3662  
Db 3619 CTAAGTCTGTAGTCTCCGAAAGATGTCAAACCAACTTTTAATACCGATGTTACGAACGG 3678  
Qy 3663 CGGCGGAACCGGCTCAGGACTACTTCGAAATCTGGTTGCAATGATTTAAAGAA 3722

Db 3679 CGGCAGAAATGCTCCCCAGACTGGACTGTGTGAAAAATCTAGTTGCGATGATTTAAAGAA 3738  
Qy 3723 ATTTCAACGCACACAGACCTGACCGGGACGATTTGACATTTGAGAGCACCGCATCTGTTGTAG 3782  
Db 3739 ATTTTAATTCACCAGAGTTGTCGGAGTAGTTGATATTGAAATACTGCACTCTTTAGTGG 3798  
Qy 3783 TAGATAAGTTTTTTGTATAGCTATTTTATTAAGAAAAAATAACAAAAAATTTGCTG 3842  
Db 3799 TAGATAAGTTTTTTGTATAGTTATTTTAAAGAAAAAAGAAACCAAAACAAAAATTTT 3858  
Qy 3843 GAGTGATGACGAAGGATTCATGATCAGATGTTTGGAAAAACAGAAAGAGTACTATTGG 3902  
Db 3859 CACTGTTTGTAGAGAGTCTCTCAATAGGTGATAGCAAGCAAGCAAGTCACAATTG 3918  
Qy 3903 ACAGCTTGGCTAACTACATTTTACAGATCTGCGGCCATTCGATCAGTACAGACATGA 3962  
Db 3919 GTCAGTTGGCCGATTTTGAATTTTGTGGATCTTCCAGCCGTTGATCAGTACAGGCATATGA 3978  
Qy 3963 TCAAGGCTCAACCAAAACAGAAATTTGACCTTTCAATTCAGATGATACCTGCTCTGC 4022  
Db 3979 TTAAGCGCAACCGAAGCAGAACTGGATCTGTCAATTCAGACAGAAATATCCAGCGTGC 4038  
Qy 4023 AACAATTTGTCTACCATTCGAACAGATCAACGGTATTTTGG--COGGTTTCTCAGAGC 4079  
Db 4039 AAGGATTTGTGATCATTTCAAGAAATCAACCAATATTTGGTCTCTTTTCAGTGAGC 4098  
Qy 4080 TTACAAGTTGCTGCTCGAGGCAATTTGATTTCTAAGAAAGTTTCTTTTCTTTTACTAGGAAA 4139  
Db 4099 TTACAAGGCAATTTACTTGACAGTATTTGACTCAAGCAGATTTCTTTGTTCTTACGAGAAA 4158  
Qy 4140 CTCAGAAACAGATTCAGAAATTTTCTCGGATCTCGACTCGCAGGTTCCTATGATGTTG 4199  
Db 4159 CACCGGCTCAGATCGAAGATTTCTTCGGAGATCTAGACAGTATGTCCTCAATGACGATC 4218  
Qy 4200 TAGAATCTGGATATTTCTTAAGTATGATAAGTCAAGAACGAGTTTCATTTGCTGTAGAGT 4259  
Db 4219 TTGAGTTGATGTTTCAAGTATGATAAGTCTCAAAACGAGTTTCATTTGCTGTGAGT 4278  
Qy 4260 ATGAAATATGAAAAAGATTGGGCTCTCAATGAGTTTTTGGCCGAAGTGTGGAACAAGGC 4319  
Db 4279 ACGAATCTCGAGGAGACTGGGCTCGAGGATTTCTTGGCAGAAAGTGTGGAACAAGGC 4338  
Qy 4320 ACAGGAAACAACTTTGAAGGATTAATCTCGAATCAAGACATGCTCTGGTATCAAA 4379  
Db 4339 ATAGAAAAACCCCTGAAAGATTACACTGCTGATATAAAACGTTTATTTGTTACCGA 4398  
Qy 4380 GGAAGCGGTGATGTGACTACTTTTCATCGGCAATCTGTTTAAATAGCAGTTCCTTGG 4439  
Db 4399 GAAAGAGTGTGATGTATCAACTTTTATCGTAAATCCGTCATCATTTGCTTCGTGCTTG 4458  
Qy 4440 GTTCAATGTTACCGATGGAAGAGGTCAATAAGGTGCTTTTGTGGAGAGATTCGCTTT 4499  
Db 4459 CATCAATGCTCCCGATGGAATAATTTGATAAAGAGGCTTCTCGGAGATGACAGTTTGT 4518  
Qy 4500 TGTATTTTCCAAGGGTTTGGATTTCCCTGACATTCAGTCAATGCTTAATCTCATGTGA 4559  
Db 4519 TGTACTTTCTTAAGGGTTGTGAGTATCCCGATATACAAACAGCTGCCAATCTAATGTGA 4578  
Qy 4560 ATTTTGAAGGCAAACTCTATAGAAAGAGGTACGGTTACTTTTGTGGTAGATACATCATAC 4619  
Db 4579 ATTTTGAAGGCAAACTGTTCAAGNAGCAATATGGGTACTTCTCGGGAGGTACGTGATTC 4638  
Qy 4620 ACCATGATAAGGGAGCAATAGTGTATTTATGATCTTTTGAAGTTGATCTCCAAACTTGGGG 4679  
Db 4639 ATCAGATAGAGGTTGATAGTATCTACGACCTTTTGAAGCTGATTTTCGAAACTTGGTG 4698  
Qy 4680 CAAAACATATCAGGATTAATGATCACTTAGAGGTTTAAAGGTCTCTTTGTCGATGTTG 4739  
Db 4699 CTAACACATCAAGGATTTGGATCATTTGGAGGAGTTTCAGAAAGATCCCTCTGTGATGTTG 4758  
Qy 4740 CTTGTTCCGTCGGAACCTGGTGTCTTAGGCTTTCCGCAAGCTGAAACGCACTATCAAGGAG 4799  
Db 4759 CTGAGTCTGTT---GAAACAATTCGCGTATTACACAAATTTGACGACGCTGTTGGGAGG 4815



```

/translation="MATQTATSSALLETVRGNTLVLNDLAKRLDYDTAVDFNARDR
RPKNVSKVSEBQTLIATKAYPEFOITFYNTQNAVHSLAGGLRSLELYLMQIPIYG
SLTYDIGNFASHLFGKRAYVHCMPNLDVDRIMRHEGOKDSIELYLSRLERGNKHPV
NFKQEAERFAEMNEVLLCHDTFOTCRHSQECYTVGVYALHSHYDIYDPADEFGAALL
RKNVHCYAAAFHSENEVLLSHDHNDEINACFORGDRLTFSFASSTLNHSHSVN
ILKVYCKTYPASNREYVMEFLVTRVNTWFCFSRDTFLFKGVAKHGVDSQEPYK
AMEDAHYKKTILAMCNERILLSDSSVNYWFPKMDVNIIVPUDISLSTSKRTRKEV
LVSKDFVYTLNIRTYOAKALTYSNVLSFVESIRSRVINGVTARSEMDVDKLLQS
IKTENALETIRVPLDYVTFHDLRVSEYKMSVDMPLDIRLAKMEETEMYNALSELVLI
KSTENALEIRVPLDYVTFHDLRVSEYKMSVDMPLDIRLAKMEETEMYNALSELVLI
KNSDKPDVDVFSQMSLEVDPMATAKVIIVAMNSESGLTLTEOQTEANVALALQDS
EKASDGLVTVSRDVEPSIRGSMARGELQLAGSDVPSESSYTRSEIESLSQFHMA
TAGSLIHKQWCSIVYTGPKVQOMKNFIDLSVLSAAVNSNLVKILKDTAIDLETRQ
KFGVLDVASRWLVKPSAKKHANGVETHARKYHVALLEHDEFIITCDNWRVAVSS
EVSUYSPMAKLRITVLDKDGEGEHPVSAKVLDVGPCKTKKILSRVNFEDLILV
PGQAASRRRANASGIIIVATKDNVTVDSFLMNGKAGQCFKRLFIDEGMLHTG
CVNLFVMSLCDIAYVGDYQOIPYINRVTFGPYPAPFAKLEVDEVETRETKLCPAD
VTHFLNQRVGHVWCTSSSEKSKSQEMVSGAASINPVSKPLKGLITFTQSDKEALLS
RGVADVHTHEVQGETYADVSLVRLTTPVSIIVARDSPHVLVLSLRSKLSKYTVVM
DPLVSIITRDLERVSYLLDMYKVDAGTQXQLOVDSVFNFLNFAAPKTDGIDSMQFY
YDKCLPGNLTLLANNYDAVTMKLTDIISLNVKDCILDMKSVAAKPKDVKPTLIPMVRTAA
EMPROTGLLENLAVMIKRNFSPELSGVDIENTASLVVDKFFDSVLLKEKRPKNF
SLPSESINLWIAKQEOVTIGOLADFDVLDPAVDYRHMIAKQPKKLDLSLOTEVP
ALQTIYVHKKINAI FGLSELTROLLSDISRSFLFTRKTPAQIETEDFGDLDSHV
FMVLEUDVSKDKSQEFHCAVEIETWRRLGLEDFLAEVKQGHRTKLUKDYTAGIK
TCLWYQRSGDVITFTIGNTVITIASCLASMLPMKLIKGAFCGDDSLLYFPKGCHEYDI
QOANLNMNFEAKLFFKQYGFYGRVYVHHDRGCIYVYDPLKILSLKLAGIKIDWPHI
EEFPRSLCDVAESLNNAYYQLDDAVGEVHTKTPGGSFVYKSLVKYLSKVLPRSLF
LDGSSC"

```

CDS

```

72..3422
/function="protein containing methyltransferase and
RNA-helicase domains"
/notes="ORF1"

```

```

/codon_start=1
/product="126K protein"
/protein_id="CAB07438.1"
/db_xref="GI:1890234"
/db_xref="GOA:P89676"

```

```

/translation="MATQTATSSALLETVRGNTLVLNDLAKRLDYDTAVDFNARDR
RPKNVSKVSEBQTLIATKAYPEFOITFYNTQNAVHSLAGGLRSLELYLMQIPIYG
SLTYDIGNFASHLFGKRAYVHCMPNLDVDRIMRHEGOKDSIELYLSRLERGNKHPV
NFKQEAERFAEMNEVLLCHDTFOTCRHSQECYTVGVYALHSHYDIYDPADEFGAALL
RKNVHCYAAAFHSENEVLLSHDHNDEINACFORGDRLTFSFASSTLNHSHSVN
ILKVYCKTYPASNREYVMEFLVTRVNTWFCFSRDTFLFKGVAKHGVDSQEPYK
AMEDAHYKKTILAMCNERILLSDSSVNYWFPKMDVNIIVPUDISLSTSKRTRKEV
LVSKDFVYTLNIRTYOAKALTYSNVLSFVESIRSRVINGVTARSEMDVDKLLQS
IKTENALETIRVPLDYVTFHDLRVSEYKMSVDMPLDIRKMEETEMYNALSELVLI
KNSDKPDVDVFSQMSLEVDPMATAKVIIVAMNSESGLTLTEOQTEANVALALQDS
EKASDGLVTVSRDVEPSIRGSMARGELQLAGSDVPSESSYTRSEIESLSQFHMA
TAGSLIHKQWCSIVYTGPKVQOMKNFIDLSVLSAAVNSNLVKILKDTAIDLETRQ
KFGVLDVASRWLVKPSAKKHANGVETHARKYHVALLEHDEFIITCDNWRVAVSS
EVSUYSPMAKLRITVLDKDGEGEHPVSAKVLDVGPCKTKKILSRVNFEDLILV
PGQAASRRRANASGIIIVATKDNVTVDSFLMNGKAGQCFKRLFIDEGMLHTG
CVNLFVMSLCDIAYVGDYQOIPYINRVTFGPYPAPFAKLEVDEVETRETKLCPAD
VTHFLNQRVGHVWCTSSSEKSKSQEMVSGAASINPVSKPLKGLITFTQSDKEALLS
RGVADVHTHEVQGETYADVSLVRLTTPVSIIVARDSPHVLVLSLRSKLSKYTVVM
DPLVSIITRDLERVSYLLDMYKVDAGTQ"
4906..5700
/function="movement protein"
/codon_start=1
/product="29K protein"
/protein_id="CAB07440.1"
/db_xref="GI:1890236"
/db_xref="GOA:P03584"
/db_xref="Swiss-Prot:P03584"
/translation="MALVVRGKVNINEFIDLSKSEKLLPSNFTPVKSVWVKDKIMV
HENESJSEVNLKGLVIEGGYVCLVGLVSGWNLDPNCRGGVSGWNRMERADE
ATLGSYYTAAKRRFQFPQVNYGITTAKDAEKNIWVLNINKVKSAGYCPILSLEFV
SVCTIVYKNNTIKLGLREKTVSVNDGGPMELSEEVDEPMENVPMSVRLKPRKTSKRG
PKNNNLGKRSGRPKPKSFDEVEKEFDNLIEDAETSIVADSDSY"
5703..6182

```

CDS

```

/function="coat protein"
/notes="ORF4"
/codon_start=1
/product="17.7K protein"
/protein_id="CAB07441.1"
/db_xref="GI:1890237"
/db_xref="GOA:P89677"
/db_xref="Swiss-Prot:P89677"
/translation="MSYPTTSPQVFLSSVWADPIELNLVNTSLGNQFQTOQARTT
VOKQSEVWKPPQSTVPFPGDVYKVRVNAVLDPLITALLGTFTDRNSIIEVENROS
PTTAETLDTRRVDATVAIRSAINNLVNLVRLGTGLYNQNTFESMSGLVWTSAPAS"

```

## ORIGIN

```

Query Match      41.2%; Score 2617.4; DB 14; Length 6383;
Best Local Similarity 64.2%; Pred. No. 0;
Matches 4091; Conservative 0; Mismatches 2231; Indels 55; Gaps 8;

Qy 18 CGACAACAACATTAACAACAAACACATATTAACAACAACAACAACAACAATGCAC 77
Db 23 CAACAACAACAACAACAACATTAACAATTC-----TACAACATACAATGCGAT 78
Qy 78 ACATACAATCTATAATTAGCAACGCCCTCTCTGAAAGCGTGAGTGTAAAAACAACCTCG 137
Db 79 ACACAACAACAGCCACATCGTCCGCTTGTGAGACCGTCCGAGGTAAACAATACCTTGG 138
Qy 138 TTAATGACCTTGCAGAAAGCGCATGTACGATACGGCCGCTGGGAAGAATTTAAACGCCGCG 197
Db 139 TCAACGATCTTGCAGAAAGCGCGTCTATATACACACAGCGGTGCGATGAATTTAATGCTAGG 198
Qy 198 ACCGTAGACCAAGAGTCAACTTTTCCAAAACATTAATGCGAGAGCAACAACGCTTCTAGTCT 257
Db 199 ACCGAGGCGCTAAAGTCAATTTTCCAAAAGTAGTAGCGAAGAACAGACGCTTATTGCAA 258
Qy 258 CCAACGCGTACCCGAGGTCCAGATTACCTTTTATAATACTCAAAATGCCGTACACAGTT 317
Db 259 CCAAAGCCTACCAGAAATTCAAATTTACATTTACACACAGCAGACGCTGTGATTTCCC 318
Qy 318 TGCTCGAGGTTTGAAGACATTAAGAATATCGATATCTGATGTCTAAGTTCCCTATGGAT 377
Db 319 TTGAGGCGGTCTCCGATCATTAGAATTTGGAATATCTGATGATGCAAAATTCCTTAOCGAT 378
Qy 378 CGCCGACATATGATATAGTGGAACTTTGCAGACATATTTGTTTCAAGGCGGAGGATACG 437
Db 379 CATTGACATATGATATCGAGGTAATTTTGCATCTCATCTGTTTCAAGGCGGAGGATACG 438
Qy 438 TGCATTGCTGTATGCCCAATCTCGACATACAGATATAATAGCGCACGAGGACGAAAGG 497
Db 439 TTCACCTGCTATGCCCAATCTGATGTCCGACACATATGCGGACAGGAGGCGCAAGG 498
Qy 498 ACTCAATTGAGATGATTTGTCCAGATTTGTCTGTTCTTAAAGGTAATTTCTTGAGTTTC 557
Db 499 ACAGTATTGAACATATACCTTTCTAGGCTCGAGAGGGGCAACAAACATGTCCCAACTTCC 558
Qy 558 AAAGGAGGCTTTTAAACAGGTATGCAAGAGCTCCCAACGAGCTCTGCTGCTCTTAAACTTT 617
Db 559 AAAAGGAAGCTTTTCAAGAGATACGCTGAAATGCAAAACGAAGTAGTCTGTGACGATACTT 618
Qy 618 TTCAGGATTTCTGAATATACATCCGCCAGAGAAATAGTGTGTAAGATACGCTGTGCTCTGC 677
Db 619 TCCAAACGCTAGGCATTCTCAAGATTTTACACGGAGAGGTGTATGCTATTGCTTTGC 678
Qy 678 ACAGTTTGTATGATATATTCCTGTGATGAGTTTGGAGCTGGTGTAAATCTTAAAGATATAC 737
Db 679 ATAGTATATACGATATACCTTCCGCGAGGTTCGCGCGGCGACCTGCTGAGAAAGATGTAC 738
Qy 738 ATGTATGTTATGCGCGCTTTCCACTTTTCCGAGAAATTTACTCTTCGAGATTTACACGTCA 797
Db 739 ATGTATGTTATGCGCGCTTTCCACTTTTCCGAGAAATTTACTCTTCGAGATTTACACGTCA 798
Qy 798 CGCTTAATGAATAGGCGCAACTTTTCAAGAGAGAGAGTGTATGATGTTTCTTTTCTTTTG 857
Db 799 ACCTCGATGAGATCAATGCAATGTTTCCAAAGAGATGAGACAGGTTTGACTTTTTTCTTTG 858

```

CDS





Db 3019 GACACGTAATGTGCACGCTCTTCTGAAAGAAATCAGTTTCCAGGAAATGTTAGTGGG 3078  
Qy 3063 AGGAGTACTAAACAGTGTCTTCAAACACCTAAAGGGAAAAATTGTAACCTTCACTCAGG 3122  
Db 3079 CTGGCTCTAATCAATCTGTCTCAAGCCGCTTAAAGGGAAAAATTGTAACCTTCACTCAGT 3138  
Qy 3123 CTGATAAATTTGAGTTAGAGGAGAAGGCTTAAAGAAATGTGAACACCGTTTCATGAGATCC 3182  
Db 3139 CTGACAGGAGGCCCTTCTCTCAAGGGGCTATGCAGATGCTCATCTGTATCATGAGGTAC 3198  
Qy 3183 AAGGAGAAACCTTTGAAGATGTGTGCTGTGTCAGATGAGCGGCACTCCACTGACTCTGA 3242  
Db 3199 AAGGTGAGCGTATGACAGCGTATCGTTAGTTCGACTTAAACACCTCAGCCCTGTATCTATCA 3258  
Qy 3243 TTTTCAAGTCTTCCCGCATGTTCTAGTCTGCTCTGACTAGACACACAAAGAGCTTCAAAAT 3302  
Db 3259 TCGCAAGAGACAGTCCGCAATGTTCTGTCTGTTGTCAAGACACACAAATCCCTAAAGT 3318  
Qy 3303 ATTACACCGTAGTGTAGATCTTTTAGTACAGATAATTAGTGAATTTGCTTCTTTAAAGCT 3362  
Db 3319 ACTACACCGTTGTGATGATGATCTTTAGTGTAGTATCAATTAGAGATTTAGAACGGGTTAGTA 3378  
Qy 3363 CTTTCTTTTAGAAATGTATGTAGACAGAGTAGTAGTAGCAATTAACAGATGATG 3422  
Db 3379 GTTACTTTATAGACATGTACAAAGTAGATGACGACTCTCAATAGCAATTAACAGTCCACT 3438  
Qy 3423 CAGTGTTCAAAGGTCATAATCTCTTGTGGCAACACCTAAATCAGGAGACTTTCCAGATC 3482  
Db 3439 CTGTGTTTAAATTTCAATCTTTTGTAGCAGCTCCAAAGACTGGAGATATATCTGATA 3498  
Qy 3483 TACAGTTCTATACGATGTATGCCCTCGTGTATAGTACTATATCTTTAAACAGTATGATG 3542  
Db 3499 TGCAATTTTACTATGATAAGTGTCTTCTCGGAACAGCAGCTTGTGTAACAACTACGACG 3558  
Qy 3543 CTGTTACATGAGGTTAGTGATTAATAGTCTTAAATGTAAGGATGTGTTCTTGATTTTT 3602  
Db 3559 CTGTTACCATGAATTTGACTGACATTTCTCTGAATGTCAAGATTTGATATAGATATGT 3618  
Qy 3603 CCAAAAGTATTTCCGATGCAAAAGGAGTGAACCATCTGTAGAGCCAGTTTGGGTACCG 3662  
Db 3619 CTAAAGTCTGTAGTGTCTCGGAAGATGTCAACCACTTTAATACCGATGGTACGAACGG 3678  
Qy 3663 CGCGGGAACCGCCAAAGGCTGCAGGACTACTCGAAATCTGTTGCAATGATTAAGAA 3722  
Db 3679 CGGCAGAAATGCTCGCCAGACTGCACTGTTGGAAATCTAGTTGCGATGATTAAGAA 3738  
Qy 3723 ATTTCAAGCACAGACTGACGGGGAGGATGATGATGAGAGCACCGCATCTGTTCTAG 3782  
Db 3739 ATTTTAATTCACAGAGTTGTCCGGAGTAGTTGATATGAAATACTGCACTCTTTAGTGG 3798  
Qy 3783 TAGATAAGTTTTTTTGTAGCTATTTTATTAAGAAAGAAATATACACAAATAATTTGCTG 3842  
Db 3799 TAGATAAGTTTTTTTGTAGTTATTTTACTTAAGGAAAGAAAGAAACCAAAATTTTT 3858  
Qy 3843 GAGTGATGACGAAGATTTCAATGATGATGATGTTGGAAACAGGAAAGAGTACTATGG 3902  
Db 3859 CACTGTTTAGTAGAGTCTCTCAATAGTGGATAGCAAGCAAGCAAGTCACAATG 3918  
Qy 3903 ACGACTTGGCTAACTACAAATTTTACAGTCTGCGGCCATCGATCAGTACAGACACATGA 3962  
Db 3919 GTCAGTTGGCCGATTTTGTACTTTTGTGATCTTCAGCGGTTGATCAGTACAGGCATATGA 3978  
Qy 3963 TCAAGGCTCAACCAAAACAGAAATTTGACCTTTTCAATTCAGAAATGATACCTGCTGTC 4022  
Db 3979 TTAAGCGCAACCGNAGCAGAACTGGATCTGTAATTCAGACAGATATCCAGGTTGC 4038  
Qy 4023 AAACAATTTGCTACCATTCGAAGCAGATCAACGGTATTTTGG---CCGGTTTCTCAGAC 4079  
Db 4039 AAACGATTTGTATCATTTCAAAGAAATCAACGCAATATTTGGTCTCTTTTCACTGAGC 4098  
Qy 4080 TTACAGGTTCTGCTCGAGGCAATTTGATTTCTAAGAGTTTCTTTTCTTTACTAGGAAA 4139  
Db 4099 TTACAGGCAATTTACTTGACAGATTTGACTCAAGCAGATTTCTTGTCTTTTACGAGAAAGA 4158

Qy 4140 CTCCAGAACAGATTCAGAAATTTTCTCGGATCTCGACTCGCAGCTTCTCTATGATGTGT 4199  
Db 4159 CACCGGCTCAGATCGAAGATTTCTTCGGAGATCTAGACAGTCACTGCTCCCAATGGAACGTAC 4218  
Qy 4200 TAGAACTGATATTTCTTAAGTATGATAGTCAAGAACGAGTTTCTTATGCTGTAGAGT 4259  
Db 4219 TTGAGTTGGATTTCTGAAAGTATGACAGTCCCANAACGAGTTTCTATGCTGTGTAGT 4278  
Qy 4260 ATCAAAATATGAAAGATTTGGGTCTCAATGAGTTTCTTGGCCGAAAGTGTGAAACAAGGGC 4319  
Db 4279 ACGAATCTGGAGAGACTGGGTCTGGAGATTTCTTGGCAGAAAGTGTGAAACACAGGGC 4338  
Qy 4320 ACAGGAAAAACAATTTGAAGGATTAATCTGCTGGAATCAAGACATGTTCTGTGTATCAAA 4379  
Db 4339 ATAGAAAAACCACTCTGAAAGATTAACACTGCTGATATAAAAACTGTTATATGTTACCA 4398  
Qy 4380 GGAAGCCGTGATGATGACTTCTTTCATCGGCAATCTGTTAATAATAGCAGCTTGTCTGG 4439  
Db 4399 GAAAGAGTGGTGTATGTTTACAACTTTTATCGTAAATACCGTCATCTGCTGCTGCTTG 4458  
Qy 4440 GTTCAATGTTTACCGATGGAAGGTCAATAAAGGTCTTTTCTGGAGAGATTTCCGTTT 4499  
Db 4459 CATCAATGCTCCGATGGAATAATGATAAAGAGCCTTCTCGGAGATGATAGTTGT 4518  
Qy 4500 TGTATTTTCCAAAGGTTTGGATTTTCCCTGACATTCAGTCATGTCATGCTCATGTTGGA 4559  
Db 4519 TGTACTTTTCTAAGGTTTGTGAGTATCCCGATATACAAAGCTGCTTAATCTAATGTGGA 4578  
Qy 4560 ATTTGAGGCAAACTGTATAGAAAGGTACGGTTACTTTTGTGGTATGATACATCATAC 4619  
Db 4579 ATTTGAGGCAAACTGTTCAAGAACCAATATGGGTACTTCTCGGAGGATGATCGTATTC 4638  
Qy 4620 ACCATGATAGGAGCAATAGTGTATTAATGATCTTTGAAAGTTGATCTCAAACTTCGGG 4679  
Db 4639 ATCAGATAGAGTTGCAATAGTATATCTACAGCCCTTTGAACTGATTTTCGAACTTCGTG 4698  
Qy 4680 CAAAAACATATCAAGGATTTATGATCACTTAGAAGATTAAAGGTGCTTTTGTGGATGTTG 4739  
Db 4699 CTAACACATCAAGGATTTGGATCAATTTGGAGGATTCAGAAAGATCCCTCTGTGATGTTG 4758  
Qy 4740 CTTGTTGCTCGGAACTGGTCTTAGGCTTTCCGAGCTTGAACGAGCTATCAAGCAGG 4799  
Db 4759 CTGAGTGGTT---GAACAATTTGCGGTATTAACACAAATTTGGAGCAGCTGTTGGGAGG 4815  
Qy 4800 TTCAAAAACCGGATTTGATGCTGTTTCTTAAATTTGTTTAAACAAATTTTGTGTG 4859  
Db 4816 TTCAAAAACCGCCCACTGTTTCTGTTTCTTAAAGATTTTAGTTAAGTATTTGTGAG 4875  
Qy 4860 ATAAATTTTATTTAGAACTTTGTTTTTAAATGCTGTTAGTCTCAGAGATACCTGCAAA 4919  
Db 4876 ATAAAGTTTGTTTAGAAATTTATTTCTTGATGCTCTAGTTGTTTAAAGGTAAGGTAAT 4935  
Qy 4920 ATTACGAGTTTCAATGATCTTTTGAACAGATGAGATCTTCGGCATTCATGACTAAG 4979  
Db 4936 ATTAATGAGTTTATCGATCTGCAAGTCTGAGAACTTCTACCGTCATGTTTCAACGCT 4995  
Qy 4980 GTCAAGAGTGTGAATATCCACTGTGCAAGATTTATGGCTGTAAAGATGATGATGTTCTT 5039  
Db 4996 GTAAAGAGTGTATGTTTCAAGGTTGATAAGATTTATGGTCCATGAAATGAATCATTTG 5055  
Qy 5040 TCTGATGTAGATTTACTTTAAAGGTTTAAAGTTAGTTAGTTAAGAAAGGTTATGTTGCTAGCT 5099  
Db 5056 TCTGAAGTAAATCTCTTTAAAGGTTTAAACCTTATAGAAGTGGGTATGTTTGTGTTAGTT 5115  
Qy 5100 GATTTGTTAGTGTCTGGGAGTGAATCTCCCGATTAACCTGCGTGTGTTGTTGCTAGTTT 5159  
Db 5116 GGTCTTGTGTTGCTGGGTGAGTGAATTTTACAGATTAATTTGCCGTGTTGTTGAGTGTG 5175  
Qy 5160 TGTATTTGATAGAGAAATCAAAAGGAGTAAAGAAAGCAACGCTGTTGGGTGCTATCAAGCC 5219  
Db 5176 TGCATGTTTGAACAGAAATGGAAGAGCGGACGACCAACCTGTTGGGTCTATATTACACT 5235



5220 CTTGCTTGCAGAAAGAAATTTCTTTTAAAGCTAATCCCTAATTTATTTCAATACATCCGAG 5279  
5236 GCTGCTGTAAAGAAAGCGGTTTCAAGTTTAAAGTGTCCCAAAATTTACGGTATTACAAAG 5295  
5280 GATGCTGAGAGACACCGCTGCGAAGTGTAGTGAATATCAAGAGAGTGGCTATGGAAGAA 5339  
5296 GATGCAGAAAGAAACATATATGCGAGGTTCTAGTAAATATTAAAAATGTAATAATGAGTGG 5355  
5340 GGATATGCTGCTTTATCTTTGGAGTTCGTTTCAATTTGTGTAGTACATATAAAATATATGTA 5399  
5356 GGCTACTGCTCTTTGTTCATTAGAAATTTGTCTGTGTATTTGTTTATATAAAATATATA 5415  
5400 AGAAAAGGTTTTCAGGGAACGTATTTTGGAGTGACAGAGCGCTCGCAATTTGAACCTACT 5459  
5416 AATTGGGTTTGGAGGAAAGTAACGAGTGTGACGATGAGGAGCCATGGAACCTTCA 5475  
5460 GAAAAGGTTTGTGAGGAGTTCGTGAGTGAAGTACCAATGCTGTGAACCTCGAAAAGGTT 5519  
5476 GAAGAGTGTGATGAGTTCATGGAGAAATGTTCCAATGTCGTTAGACTCGCAAGTTT 5535  
5520 CCGGAAACAAAGAAATGGTAGTAATAATGTTTAATAAT----- 5561  
5536 CGAACCAAAATCCTCAAAAGAGGTCGCAAAATAATAATAATTTAGGTAAAGGCGCTTCA 5595  
5562 ---AAGAAATAAATACAGTGTGTAGAGGTTTAAATTTGAGGAAATTTGAGGATAAT 5618  
5596 GCGGAGGCGCTTAAACCAAAAGGTTTGAAGTTGAAAAGAGTTTGATTAATTTGATT 5655  
5619 GTAAGTGTGACGAGTCTATCGGCTCATCGAGTACGTTTAAATCAATATGCTTATACAA 5678  
5656 GAAGATGAAGCGAGAGCTCGTTCGGGATCTGATTCTGTTAATATATGTTTACCAG 5715  
5679 TCAATCTCCGAGCCAAATTTGTTTACTTATCTTCGCTTACGAGATCTGTCAGCTGA 5738  
5716 TCACTTCTCCATCGCAATTTGTTTGTGTCATCTGTATGGGCTGACCTATAGAAATGTT 5775  
5739 TCATCTGTGTACAAATGATTTGGGTAAACGAGTTTCAACGCAACAGCTAGGACAAAG 5798  
5776 TAAACGTTTGTACAAATCTGTTAGGTAAACGAGTTTCAACGCAACAGCAAGCAACTACTG 5835  
5799 TCAACAGCAATTTGCGGATGCTGCAACCTGCTGCTAGTATGACAGTGAAGTTTCTG 5858  
5836 TTTCAAGCAGTTTACGAGGTTGGAACCTTTTCCCTCAGACCACTGATGATTTCTG 5895  
5859 CATCGGATTTCTATGTGTATAGATATAATTCGAGCTTGCATCGTTGATCAAGCGCTTAT 5918  
5896 GCGATGTTTATAAGGTGTACAGTACAATGTCAGTCTTTTATGATCTTTTAACTGCTGCTG 5955  
5919 TAAATAGCTTTGATCTAGAAATAGAAATATAGAGGTTGTAATCAACCGGACCGCAATA 5978  
5956 TGGGGACTTTTCGATCTAGGAATAGCAATAATCGAAGTAGAAATCGGCAAGTCCGACAA 6015  
5979 CTACTGAAATCGTTAAGCGGACTCAGAGGTTAGACGATGCTAGTGTAGCTATAAGGGCTT 6038  
6016 CAGCTGAAGCTTATGATGCTACCGGAGGTTAGACGATGATGATGATGATGATGATGATG 6075  
6039 CAATCAATAATTTGGCTTAATGAATCTGTTTGTGGAATCTGCAATGTTTCAATCAAGCAGCT 6098  
6076 CTATAATAATTTTATGTTAATGAATAGTAAAGGTTAGTGAAGTCTGAGCTGTACAAATCAG 6135  
6099 TTGAGAGCTGTAGTGAATGCTGACGACCAACTCGGCTACTTACTGATTTGTTGTTGAG 6158  
6136 TTGAAAGTATGCTGCGGTTTGTGCGACCTCTGACCTGATCTGATCTGATCTGATCTGAT 6195  
6159 ATTTCCTAAATAAAGTTCGCTGAAGACTTAAATTTCAAGGTTGCTGATACCAAAATCAG 6218  
6196 GAAATATAAA-----GTTTGTGTTTCTAAACACACGTTGATGATGATGATGATGATG 6247  
6219 AGTGGTTGTTTGTGCTCACTTAAATATATACGATTTGATCTGATCTGATCTGATCTGATCTG 6278  
6248 AGTGTGTTTCCCTCCACTTAAATCGAAGGTTAGTGTCTTGGAGCGCGCGGAGTAAACATA 6307  
6279 TGTGATGGTGTATGCTGTGTTATGCGGTAAACATCGGAGAGTTTGAATCTCTCCCTTAA 6338

Db 6308 TATGCT-TCATATATGTCGTAGGACGCTAAAAAGCGAGGATTCGAATTTCCCGCGAA 6366  
QY 6339 CCGCCGGTAGCGGCCA 6355  
Db 6367 CCGCCGGTTGCGGCCA 6383  
RESULT 13  
TMO417701 6383 bp RNA linear VRL 31-OCT-2001  
Tomato mosaic virus camellia strain, complete genome.  
LOCUS AJ417701  
DEFINITION AJ417701.1 GI:16580717  
VERSION coat protein; complete genome; movement protein gene; RNA-depended  
KEYWORDS RNA polymerase.  
SOURCE Tomato mosaic virus  
ORGANISM Tomato mosaic virus  
VIRUSES; serNA positive-strand viruses, no DNA stage; Tobamovirus.  
REFERENCE 1  
AUTHORS Xue, C., Tao, X. and Zhou, X.  
TITLE Molecular aspect of pathogenicity differences on Nicotiana tabacum  
induced by Tobacco mosaic virus and Tomato mosaic virus  
JOURNAL Unpublished  
REFERENCE 2  
AUTHORS Zhou, X. and Xue, C.  
TITLE Characterization and complete nucleotide sequence of two isolates  
of tomato mosaic virus  
JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 6383)  
AUTHORS Zhou, X.  
TITLE Direct Submission  
JOURNAL Submitted (26-OCT-2001) Zhou X., Zhejiang University, Institute of  
Biotechnology, Kaixuan Road 268, Hangzhou, Zhejiang, 310029, CHINA  
FEATURES  
1. 6383  
/organism="Tomato mosaic virus"  
/virus="Tobamovirus"  
/mol\_type="genomic RNA"  
/strain="camellia"  
/db\_xref="taxon:12253"  
/country="China"  
2. 4922  
/note="contains readthrough stop codon"  
/codon\_start=1  
/transl\_except=(pos:3420..3422,aa:Gln)  
/product="RNA-depended RNA polymerase"  
/protein\_id="CAD10425.1"  
/db\_xref="GI:16580718"  
/db\_xref="GOA:Q911R9"  
/translation="MAYTQTATSSALLETVRGNNTLVNDLAKRRLDYDTAVDFNARDR  
RPKNFVKVSEOTLIATKAYPEQITPTNTQNAVHSLAGLSLELEYLAWQIPYG  
SLTYDIGNFASHLPKRAYVHCMPNLDVRIMRHEGKQDSIELYLSLRGKHVP  
NPQKAPRYAEMPEVNEVCHDTPTQCGHSECTGRVVALHLSYDIADPAGFALL  
RKNVCAAFHSEINLLEDHVNDEINACTQKDRDLTSPASESTLNTSHSVN  
ILKYKTYTPASNEVYKMFLEVRVNTWFCFKSRIDTFLLYKGVHAGVDSQEPYK  
AMEDAWHYKTTLAMCNSEIRLLEDSSVNTWFPKRDVVIINGVTSRSEWVDSKLSQ  
LVSRDVTVAHLNHTYQKALYSNLSFVBSIRSRVINGVTSRSEWVDSKLSQ  
LSMTFLATKLAVKDDLLISKFALGPKTVSOHVDEISLAFGNAPPSIKERLNRKL  
IKITENALBIRVPDLYVTFHDLVSEYKMSVMDPVLDIRKMEETEMNALSLSVL  
KNSDFVDVFSQMSLEVDPMATAKVIIVMSNESGLTLTFEQTEANVALALQDS  
EKASDGLVTSRDEVEISIKGSMARGELQAGLSDVPESSTYRNEEISLEQFMA  
TASLIHQMSIYVTGLPKVQMKNFIDSLVASLSAAVSNLVKILKDTAAIDLETRQ  
KGVLDVASRKLVPKSNHAWGVTHARKYHVALLHDEFGIITCNWRVAVSS  
ESVYDWMKLTALKKLDGEPHYSSAKVLDVGVGCGKTEILSRVNFREDLTV  
PQQAEMIRRRANASGIIVATKNVRAVDSFLMNYGKARQCFELFIDEGLMLHTG  
CVNFEVMSLCDIAVYGTQIIPYINRVTPPFAHFAKLEVEETRTTLURCPAD  
ATHFLNQRVEGVMCTSEKSKSQMSVSGAASINPVSPLKEILTFTQSKDEALLS  
RGYADVTHVEVQGTADVSLVRLTPTPVSIIAGDSPHLVSLSRHTKSLKYTYVM  
DPLVSIIRDLERSVYLLDMYKVDAGTQQLQVDSVFNKFNLPVAAKPTGDSIMQFY  
YDKCLPSTLLNNDYATMKLTDIILNVKDCILDMSKSAAPKVKPTLIEMVTA  
EMPROTGLENLVAMIKRNFNPELSGVVDIENTASLVVDKFDSDLSLLKPKPKNF



1278 TTAAATCATATCAGAAACATATCAAGCCAAAGCGTTACTTACCAGACGTTATTTCTTTCG 1337  
1279 TAAATCAATTCGTACGTACACGCGCAAGCGCTTACTTACTCCAAAGCTTTATCTTTTCG 1338  
1338 TGGAGTCTATAAGATCCCGCGTGATATCAATGGTGTACTGCTAGGTCTGGAATGGGATG 1397  
1339 TCGAATCGATTCTGTCAGAGATGATCAATTAACGGGGTTACTGCTAGGTCTGAGTGGGATG 1398  
1398 TAGATAAAGCAATCTTCAACCTTGTCAATGATCTTCTTCTTGGCAGACTAAGCTGGCTG 1457  
1399 TCGATAAATCATTTATACAGTCTTGTGATGAGCGTTCTTCTTCTACATACCAGCTTGCCG 1458  
1458 CGCTTCAAGACGATATAGTAATGGGAAGTTTCGGTCTGGATATAGACCACCTTCTGAC 1517  
1459 TCTGAAGAACGATCTTTTGAATAGCAAGTTTGACCTTGGTCCAAAACCTGTCTCAAC 1518  
1518 TTATTTGGGATGAGTGGGCAAAATTTTTTGGAAACGTTTTTCCCACTATCAAAAGAGAT 1577  
1519 ATGTGTGGATGAGATTTCCTAGCTTTCGCAATGCTTTCCCATCGATCAAGGAAGAT 1578  
1578 TGGTAGCAGGAAATTTCTGGATGTAGTGAGAAATGCTCTGAAAGATCAAGATCCAGATC 1637  
1579 TGATAAACCGGAAACTGATCAAAATTTACCGGAGAAATGCGTTAGAGATCAGGGTCCCGATC 1638  
1638 TGTATGTCACATGGAAGACAGGTTCTGTAGCTGATACCAACCAAGTCTGAGGAGTTACCGC 1697  
1639 TTTATGTCACTTTCCATGATAGGTTAGTTTCTGAGTACAAAATGTGAGTGGAATGCCGG 1698  
1698 ATCTAGATATCAAGAGGAGCTTAGAAGAACTGAGCAAAATGTACGACCGCTTATCAGAA 1757  
1699 TGCTAGACATTAGGAAGAGATGGAAGAACTGAGGAAATGTACAAATGTGCTCCGATC 1758  
1758 TATCTATCTTAAAGGTTGTGATAATTTTCAATATGCGGAAGTTTCAAGACATGTGCAAGG 1817  
1759 TGTCTGTACTTAAAAATTCAGACAAGTTTCGATGTGATGTTTTCACAGATGTGCCAGT 1818  
1818 CTTTAGATGTAGTCTGTAGTGGCAGCAGATTAATCGTTGAGTGGCCGCGAATAGAA 1877  
1819 CTTTAGAAGTCGATCCAAATGACTGCAAGCAAGGTAATAGTAGCAGTTATGAGCAACGAGA 1878  
1878 GCGGTTTAACTCTTACTTTTGAAGCCAAACCGAGGAGATGTGGCTTAAGGCTTTAAA - 1936  
1879 GTGGTCTTACTCTACAGTTTGAACAGCCACCGAGGCTAATGTGCGTAGCATTCGAC 1938  
1937 -----AGCACGGGTCTGAGCCGTGGTATGTCTTGAAACCGACATCCGAAGAGTGAACG 1991  
1939 ATTCTGAAAAGGCTTCTGATGGGCGTTGGTAGTTACTCAAGAGATGTTGAGAACTGT 1998  
1992 TAAATAAATTTCTATTGCTGAGAAAGGAGATTCCTGTGTGCGAAGATCATGGTT 2051  
1999 CCATAAAGGGTTCGATGGCCCGTGGTGAAGTTGCAATTTGGCCGATATCTGGCGAGCTTC 2058  
2052 TGACGAATGCTAACTTAGACACACAGAGTTGGAGTCCCTCAACGATTTCCATAAGGCTT 2111  
2059 CTGAACTTCAACACTAGAAACGAGGAGATTGAGTCTCTCGAGCAGTTTCATATGGCAA 2118  
2112 GCGTGGATGTGATTPACAAAGCAATGCGATCGGTTGTCTACACTGGCTCTCACTCAAAG 2171  
2119 CAGCTAGTTCTGTTAAATCAATAGCAGATGTTCGATCGGTGTACACGGGCGCTTTAAAG 2178  
2172 TTCAACAAATGAAGACTATGTGGACAGTTTGGCAGCTTGGTGTGCGCCACTGTATCAA 2231  
2179 TTCAACAAATGAAGAACTTTATAGACAGCTGGTAGCTCGCTCTGTGCTGGGTTGCGA 2238  
2232 ATCTATGCAAGTCACTAAAGATGAAGTCCGGGTATGATTTCTGATTCAGGGGAGAAAGTTG 2291  
2239 ATCTAGTGAAGATCTTAAAGATACAGCCGGATTGACCTTGAACTCTGCTCAAAAGTTTCG 2298  
2292 GTGTTTGGGATGTCACTTTTGAAGAGTGGCTCTCAAACTCGCGCCAAAGGTCATTCAT 2351  
2299 GAGTTCTGGATGTGCTTCGAAAAGGTGGCTAGTTAAACCATCGCAAGAAACCATGCAT 2358  
2352 GGGAGTTGTCTGGATTACAGGGGAAATGTTTACTGCACCTTCTATCTATTAAGGAG 2411

2359 GGGGGGTTGTTGAGACTCATGCGAGGAATATACAGTCCGATTTACTTGGACACGATGAAT 2418  
2412 ATAGAAATGGTGTACTGAGACGACTGGAGGAGGGTGGCTGTATCATCTGATACAAATGGTAT 2471  
2419 TTGGCATTTATCAGTCCGATTAACCTGGCGACGGGTGGCTGTGAGTCTTCTGAGTCGGTAGTAT 2478  
2472 ATTCTGATATTGCAAGCTCCAAAATCTGAGGAAACAAATGAGAGACGGTGAACCCACG 2531  
2479 ATTCTGATATGGCTAAACTCAGGACTCTGAAAAGATTGCTTAAAGATGGAGAACCAACG 2538  
2532 AACCTACTGCAAGATGTGTACTTGTGATGGGGTGGCTGTGTTGTTGAAAGTACAAAGGAG 2591  
2539 TTAGTTTCAGCAAAAGTGGTTTTGGTGGATGGCGTTCCAGGGTCCGGGAAACAAAGGAAA 2598  
2592 ATTTTGAAGAATTTGATCTTGTATGAGGATTTGATCTTGGTTCTGTGAAAACAAAGTGGCTG 2651  
2599 TTCTTTTCGAGAGTTAATTTTGAAGAAGATCTAATTTCTGTCCCTGGTCCGTCAGAGCTGCCG 2658  
2652 CTATGATCAGAGAGGCTTAATTCATCTGGACTGATAGAGCCCAATGGAACAAATGTGA 2711  
2659 AGATGATCAGAGAGAGCTAATGCGTCCGGGCATTAATAGTGGCTCAAAAGGATAATGTGC 2718  
2712 GAACGGTAGATTCACTTTCTAATGCATCCAAAACCGCGATCAC-----ACAAGAGGC 2762  
2719 GCGCCGTCAATTCATTTCTTGTGTAATTAACGGNAAGGGGACGCTGTCAAGTTCAAAAGAT 2778  
2763 TTTTATTTGATGAAGGTTGATGCTGCACACCGGTTGTGTAACTTCTTGGTGGCTTATCT 2822  
2779 TGTTCATAGACGAAGGTTTGTGCTGCATCTGCTGTGTTGTGTGAATTTCTTGGTTGAAATGT 2838  
2823 CTGGTTGCCACATCCGATACATTTTACGGAGATACACAGCAGATTCCTTTTCATTAACAGAG 2882  
2839 CTCTGTGCGATATGTCATATGTTTATGAGACACCAACAAATTTCCGTACATCAACAGAG 2898  
2883 TTCAGAAATTTCCGCTATCCCAACATTTTGAAGAAGTGCAGGTGGATGAAGTTGAGATGA 2942  
2899 TAACTGGTTTCCGCTACCTTCACACTTTTGCAAAATTTGAGGTCGAGAAAGTTCGAAACAA 2958  
2943 GGAGGACCACTCAGATGCCAGGTGATGTGTAATTTTTTCTTCAATCGAAGTACGAAG 3002  
2959 GAAGAACTACTCTTCTGTTGCGGCTGATGCCACACATTTCTTAAATCAAAGGTATGAAG 3018  
3003 GAGCGGTGACCAACCACTTCAACTGTACACAGATCGGTCTCATCTGAGATGATAGGCGGTA 3062  
3019 GACACGTAATGTGCACTCTTCTGAAAAGAAATCAGTTTCCAGGAAATGTTAGTGGGG 3078  
3063 AGGAGTACTTAAACAGTGTTCCAAACCACTAAAAGGGAATTTGTAACCTTTCACCTCAGG 3122  
3079 CTGCGTCTATCAATCTCTGTGTCGAGCGCTTAAAGGGGAAATTTTGACTTTTCAACAGT 3138  
3123 CTGATAAATTTTGAAGTTAGAGGAGAGGCTTATAAGAAATGTGAACACCGTTTCATGAGATCC 3182  
3139 CTGACAAAGGAGGCCCTTCTCTCAAGGGCTATGACAGATGTCCATCTGTACATGAGGTAC 3198  
3183 AGGAGAAACCTTTGAAGATGTGCTCGTGTGAGTTGAGCGCACTCCCACTGACTCTGA 3242  
3199 AAGGTGAGACTTATGACAGCAGTATCGTTAGTTCCACTAACACCTACGCTGTATCTATCA 3258  
3243 TTTTCCAAAGTCTTCCCGCAGTGTTCCTAGTCTGCTGACTAGACACACAAGAGGTTCAAAAT 3302  
3259 TCGAGGAGACAGTCCGCAATGTTCTGGTCTGTTGTCAAGACACACAAATTCCTTAAAGT 3318  
3303 ATTACACCGGTAGTGTAGATTCCTTTTAGTACAGATAATTAGTGAATTTGTCTTCTTTAAGCT 3362  
3319 ACTACACCGTTGTGATGGATCTTTTAGTTAGTATCATAGAGATTTAGAACGGGTAGTA 3378  
3363 CTTTCTCTTTTAGAAATGTATATGTTAGAGACAGGTAGTAGTAAGCAATTAACAGATGGATG 3422  
3379 GTTACTTTATAGACATGTCAAAAGTAGTGCAGGTACTCAATAGCAATTTACAGGTGCACT 3438  
3423 CAGTGTTCAAAGGTCATTAATCTTCTTGTGCGACACCTTAATCAGGAGACTTTTCCAGATC 3482

Db 3439 CTGTGTTTAAATAATTTCAATCTTTTGTAGCACTCCAAAGACTGGAGATATATCTGATA 3498  
QY 3483 TACAGTCTTATACGATGATGCTCCCTGGTAAATAGTACTATATCTTAACAGATATGATG 3542  
Db 3499 TGCATTTTACTATGATAGTGTCTTCTGGGAACAGCACGTTGTGTGAACAACTACGACG 3558  
QY 3543 CTGTACCAATGAGGTTACGATGATAATAGTCTTAAATGTGAAGGATTTGTCTTGAATTTT 3602  
Db 3559 CTGTACCAATGAAATTTGACTGACATTTCTCTGAATGTCAAAGATTCATATAGATATGT 3618  
QY 3603 CCAAAAGTATTCGGATGCGAAAGGAGTGAAACCAATGTCTAGAGCCAGTTTTCGTACCG 3662  
Db 3619 CTAAGTCTGTAGTCTGCTCGAAAGATGTCAAACCAACTTTAATACCGATGGTACGAACGG 3678  
QY 3663 CGCGCGAACCGCAAGGGCTGCAGACTACTCGAAATCTGTTTCAATGATTTAAAGAA 3722  
Db 3679 CGCGAGAAATGCTCGCCAGACTGACCTGTGTGAAATCTAGTTCCGATGATTTAAAGAA 3738  
QY 3723 ATTTCAACGACACAGACCTGACGGGACGATTTGACATTTGAGAGCACCGCATCTGTGTAG 3782  
Db 3739 ATTTTAAATTCACAGAGTGTCCGAGTAGTTGATATTTGAAATACTGCATCTTTAGTGG 3798  
QY 3783 TAGATAAGTTTTTTCATAGCTATTTTATTAAGAAAGAAATATACACAAAAATATTCGCTG 3842  
Db 3799 TAGATAAGTTTTTTCATAGTATTTTACTTAAAGGAAAGAAAGAAACCAACAAAAATTTT 3858  
QY 3843 GAGTGATGACGAAGGATTCATATGATGAGATGGTTGGAACACAGGAAGAACTACTATTGG 3902  
Db 3859 CACTGTTTAGTAGAGAGTCTCTCAATAGTGGATGACAAAGCAAGACAGTCAATTTG 3918  
QY 3903 AGCACTTGGCTAACTACAAATTTTACAGATCTGCGGCCATCGATCAGTACAGCACATGA 3962  
Db 3919 GTCAAGTGGCCGATTTTGAATTTGTGGATCTTCAGCGCGTTGATCAGTACAGGCATATGA 3978  
QY 3963 TCAAGGCTCAACAAAAACAGAAATTTGACCTTCAATTTCAAGATTTCTTCTTACTAGGAAAA 4022  
Db 3979 TTAAAGCCACCGAGCAGAACTGGATCTGTCATTTTCAACAGAGATATCCAGCGTTGC 4038  
QY 4023 AAACAATTTGTCTACCAITTCGAAGCAGATCAACGGTATTTTGG---CGGTTTCTCAGAGC 4079  
Db 4039 AAACGATTTGTATCATTTCAAGAAATTCACGCAATATTTGGTCTCTTTTCACTGAGC 4098  
QY 4080 TTCAAGGTTGCTGCTCGAGCATTTGATTTCTAAGAGTTTCTTTTCTTACTAGGAAAA 4139  
Db 4099 TTCAAGGCAATTTACTTGACAGTATTTGACTCAACGAGATTTCTTTTCTTCAAGGAAAGA 4158  
QY 4140 CTCCAGAAACAGATTCAGAAATTTTCTCGGATCTCGACTCGACGTTTCTTATGATGCTGT 4199  
Db 4159 CACCGCTCAGATCGAAGATTTCTTCGGAGATCTAGACAGTCAATGTCCTCAATGGACGTAC 4218  
QY 4200 TAGAACTGGATTTTCTAAGTATGATAGTACAGAACGAGTTTTCATTTGCTGTAGAGT 4259  
Db 4219 TTGAGTTGGATTTTTCGAAGTATGATAAGTCTCAAAACGAGTTTCAATTTGCTGTGAGT 4278  
QY 4260 ATGAAATATGAAAGATTTGGTCTCAATGAGTTTTCGGCGAGTGTGGAAACAGAGGC 4319  
Db 4279 ACGAATCTGAGGAGACTAGGTTGAGAGGATTTCTTGGCAGAAAGTGTGGAAACAGAGGC 4338  
QY 4320 ACAGAAACAACTTTTGAAGGATTTACATTTGCTGGAATCAAGACATGTCGTGGTATCAAA 4379  
Db 4339 ATAGAAACCACTCTGAAGATTTACATGCTGGTATATAAAACGTTTATTTGTTACAGA 4398  
QY 4380 GGAAGCGGTGATGTACTACTTTTCATCGCAATCTGTTATTAATAGCAGTGTGCTGG 4439  
Db 4399 GAAAGAGTGGTATGTTTACAACTTTTATCGGTATACCGTCATCATTTGCTGTCTTG 4458  
QY 4440 GTTCAATGTTTACCGATGGAAGGTCATAAAGGTGCTTTTGTGGAGAGATTTCCGTTT 4499  
Db 4459 CATCAATGCTCCCGATGGAATAATTTGATAAGAGAGCTTCTCGCGAGATGACAGTTGT 4518  
QY 4500 TGTATTTTCCAAAGGTTTGGATTTCCCTGACATTCAGTCAATGCTGCTAAATCTCATGTGGA 4559  
Db 4519 TGTACTTTCTAAGGGTTGTGAGTATCCCGATATACCAACAGAGTCTGCTAACTAATGTGA 4578

QY 4560 ATTTTGAGGCCAACTGTATAGAAAAGAGGTACGGTACTCTTTTGTGGTATGATACATCATAC 4619  
Db 4579 ATTTTGAGGCCAACTGTTCAAAGAACAAATATGGTACTTTCTCGGGAGGTACGTGATTC 4638  
QY 4620 ACCATGATAAGGGAGCAATAGTGTATTATGATCTCTTTGAAAGTTGATCTCCAACTTCGGG 4679  
Db 4639 ATCAAGATAGAGGTTGCAATAGTATATCAACGCCCTTTGAAAGCTGATTTCCAACTTCGGT 4698  
QY 4680 CAAACAATATCAAGGATTTATGATCACTTTAGAAGAGTTTAAAGGTTGTCTTTTGTGCGATTTG 4739  
Db 4699 CTAACACATCAAGGATTTGGATCATTTGGAGGATTCAGAAGATCCCTCTGTGATTTG 4758  
QY 4740 CTTGTTGCTCGGAACCTGCTAGGCTTTCCGACGCTGAACGAGCTATCAAGGAGG 4799  
Db 4759 CTGAGTCGTT---GAACAATTCGCGTATTTACACAAATTTGGACGAGCTGTTGGGAGG 4815  
QY 4800 TTCTAAACCGGATTTGATGCTGTTCTTTTAAATTTGTTTAACTTTTAACTTTTGTGTG 4859  
Db 4816 TTCTAAACCGGCCACCTGTTCTGTTTGTATTAAGAGTTTAGTTAAGTATTTGTCTAG 4875  
QY 4860 ATAAATTTTATTTAGAACTTTGTTTAAATGGCTGTTAGTCTCAGAGATCTGTCAAA 4919  
Db 4876 ATAAAGTTTGTTTAGAAAGTTTATTTCTGATGGCTTAGTTCTTAAAGGTAAGGTAAT 4935  
QY 4920 ATTAGCGAGTTCAATGATCTTTCCGAACAGGATGAGATCTTCCGGCATTTCACTGCTAAG 4979  
Db 4936 ATTAATGAGTTTATCGATCTGTCAAAGCTGAGAACTTCTCCCGCTGATGTTTCCGCT 4995  
QY 4980 GTCAAGAGTTTAAATATCGACTGTGGACAAAGATTTAGCTTTAAGAAATGATGATCTTT 5039  
Db 4996 GTAAAGAGTTTATGTTTCAAAGTTTGAATAAGATTTATGTCATGAAATGAATCATTTG 5055  
QY 5040 TCTGATGTAGATTTACTTAAAGGTTTAAAGTTAGTTTAAAGAGGATGTTGCTGCTAGCT 5099  
Db 5056 TCTGAAGTAAATCTCTTAAAGGTTTAAACCTTATAGAAGGTGGATATGTTTGTCTTAGTT 5115  
QY 5100 GATTTGCTAGTGTCTGGGAGTGGAAATCTCCGGATAACTCCCGTGGTGTGTCAGTGT 5159  
Db 5116 GGTCTTGTGTTGTCGCGTGAATTTTACAGATTAATTTGCGTGGTGTGAGTGTCT 5175  
QY 5160 TGTATTGTAGATAAGAAATGAAGAGATGAGAAACACGCTGGTGGTATCAAGCC 5219  
Db 5176 TGCAATGTTTACAAAGAAATGGAAGAGCGGACAAACCCACATGCGGTCTATATACAT 5235  
QY 5220 CTTGCTTGCAGAAAGAAATTTTCTTTAAAGCTAACTCCCTAAATTTTCAATAACATCCGAG 5279  
Db 5236 GCTGCTGTAAAGCGGTTTCAAGTTTAAAGTGGTCCCAATTTACGGTATTTACAAAG 5295  
QY 5280 GATGCTGAGAAGCACCCGTGGCAAGTGTGTAATATCAAAAGGAGTGGCTATGGAAGAA 5339  
Db 5296 GATGCAGAAAAGAACATATGCGAGGTTCTTAGTAAATATTTAAATACTAATAATGAGTGG 5355  
QY 5340 GGNATCTGCTTTATCTTTGGAGTTCGTTTCAATTTGTGTAGTACATAAAAATAATGTA 5399  
Db 5356 GGTACTGCTTCTTGTTCATTAGAAATTTGTCTGTGTATTTTATATAAATAATATA 5415  
QY 5400 AGAAAAGTTTTCAGGGAACGTTATTTTGTGTGACAGAGCGCTCGCAATTCGACTCACT 5459  
Db 5416 AATTTGGGTTTGGAGGAGAAAGTAACGATGTGAACGATGGAGGCCATGGAACCTTCA 5475  
QY 5460 GAAAAGTTTGTGAGGAGTTCTGGATGAAGTACCAATGGCTGTGAAAACCTCGAAAAGGTT 5519  
Db 5476 GAAGAAGTTTGTGATGAGTTTCATGGAGATGTTTCAATGCTCGGTTAGACTCGCAAGTTT 5535  
QY 5520 CCGGAAAAACAAAAGAAATGTTAGTATATGTTTAAAT----- 5561  
Db 5536 CGAACCAATCTCRAAAAAGAGGTCGAAAAATAATAATTTAGGTAAAGGGGCTTCA 5595  
QY 5562 ---AAGAAATAATAACAGTGTGAAGGCTTTTAAATTTGAGGAAATTCAGGATAT 5618  
Db 5596 GCGGAAGSCCTAAACCAAAAAGTTTGTGATGAAGTTGGAAAGAGTTTGTATTAATTTGAT 5655

QY 5619 GTAAGTATGACGAGTCTATCGCGTCATCGAGTACGTTTAAATCAATATGCTTATACAA 5678  
 Db 5656 GAAGATGAAGCCGAGAGCTCGGTCGCGGATCTGATTCGTAATTAATATATGCTTACTCAA 5715  
 QY 5679 TCAACTCTCCGAGCAATTTGTTTACTTATCTTCGCTTACGCGATCCTGTGCGAGCTGA 5738  
 Db 5716 TCACTTCTCATCGCAATTTGTTTGTGTCATCTGTATGGGCTGACCTATAGATTGT 5775  
 QY 5739 TCAATCTGTGTACAAATGCAATCGGTAAACGAGTTTCAACGCAACGAGCTAGGACACAG 5798  
 Db 5776 TAAACGTTTGTACAAATTCGTTAGGTAAACGAGTTTCAACGCAACGAGCAAGCAACTACTG 5835  
 QY 5799 TCAACGAGCAATTTGCGGATCGCTGGAACCTGCTAGTATGACAGTACGATTTCCCTG 5858  
 Db 5836 TCAACGAGCAATTTGCGGATCGCTGGAACCTTTCCTCAGACGCGTACGATTTCCCTG 5895  
 QY 5859 CATCGAATTTCTATGTGTATAGATATAATTCGACGCTTGATCTCGGTGATCAACGCGCTTAT 5918  
 Db 5896 GCGATGTTTATAGGTGTACAGGTACATGCTAGTTTGTAGATCTCTAATTTACTGCGTTCG 5955  
 QY 5919 TAAATAGCTTTGTATCTAGAAATAGAAATAGAGTTGATATCAACCCGCAACCGAATA 5978  
 Db 5956 TGGGCTCTTTGTATCTAGGAATAGAAATAGAAATAGAAATAGAAATAGAAATAGAAATAG 6015  
 QY 5979 CTACTGAATTCGTTAAGCGCACTCAGAGGTAGACGATGCTAGTATGATTAAGGCTT 6038  
 Db 6016 CAGCTGAACGTTAGATGCTACCGCGAGGTAGACGATGCTAGTATGATTAAGGCTT 6075  
 QY 6039 CAATCAATATTTGGCTAATGAATGCTGCTGCACTGCTGATTTCAATCAAGCAGGCT 6098  
 Db 6076 CTATCAATATTTAGTTTAACTAGTAGTAGAGTACTGACATACATCAAAATACTT 6135  
 QY 6099 TTGAGACTGCTAGTGGACTTGTCTGGACCAACTCCGGCTACTAGCTATTTGTGTAG 6158  
 Db 6136 TTGAAAGTATGCTGGGTTGCTGCGACCTCTGCACCTGCATCTTAAATGCATAGGTGCT 6195  
 QY 6159 ATTTCCTAAATGAAGCTGCTAGACTTAAATTCAGGCTGCTGATACCAATCAGC 6218  
 Db 6196 GAAATATAAA-----GTTTGTGTTTCTTAAACACACAGCTGCTAGTATGATTAAGCTATC 6247  
 QY 6219 AGTGGTGTGTTGCTCCACTTAAATATACGATTTGTCATATCTGATCCCAACAGTTTAAACCA 6278  
 Db 6248 AGTGTGTTTCCCTCCACTTAAATTCGAAGGTAGTGTCTTGAGGCGCGGAGTAAACATA 6307  
 QY 6279 TGTGATGGTGTACTGTGATGGGTATGCGGTAAACATCGGAGAGTTTGAATTCCTCCCTAA 6338  
 Db 6308 TATGGT-TCATATATGTCGCTAGGACGAGTAAAGGAGGAGTTCGAATTTCCCGCGAA 6366  
 QY 6339 CCGCCGTAGCGGCCA 6355  
 Db 6367 CCCCCGTTGGGGCCCA 6383

## RESULT 14

AB069853  
 LOCUS 6357 bp RNA linear VRL 13-FEB-2002  
 DEFINITION Pepper mild mottle virus gene for 126 kDa protein, complete cds.  
 ACCESSION AB069853  
 VERSION AB069853.1 GI:18652912  
 KEYWORDS Pepper mild mottle virus  
 SOURCE Pepper mild mottle virus  
 ORGANISM Viruses; serRNA positive-strand viruses, no DNA stage; Tobamovirus.  
 REFERENCE 1  
 Hagiwara, K., Ichiki, T. U., Ogawa, Y., Omura, T. and Tsuda, S.  
 A single amino acid substitution in 126-kDa protein of Pepper mild mottle virus associates with symptom attenuation in pepper; the complete nucleotide sequence of an attenuated strain, C-1421  
 Arch. Virol. (2002) In press  
 REFERENCE 2 (bases 1 to 6357)  
 Hagiwara, K., Ichiki, T., Ogawa, Y., Omura, T. and Tsuda, S.  
 Direct Submission  
 TITLE Submitted (10-AUG-2001) Kyoji Hagiwara, National Agricultural

Research Center, Laboratory of Virology; Kannondai, Teukuba, Ibaraki  
 305-8666, Japan (E-mail: kyoji@affrc.go.jp, Tel: 81-298-38-7030,  
 Fax: 81-298-38-8929)

## FEATURES

source Location/Qualifiers  
 1..6357  
 /organism="Pepper mild mottle virus"  
 /mol\_type="genomic RNA"  
 /isolate="C-1421"  
 /db\_xref="taxon:12239"  
 70..3423  
 /codon\_start=1  
 /product="126 kDa protein"  
 /protein\_id="BAB84693.1"  
 /db\_xref="GI:18652913"  
 /translation="MAYTQATNMAALASTLRGNPLVNDLANRRLVESAVEQCNANDR  
 RKVNFRLSISEEQTLIATKYPEFOITFYNTQNAVHSLAGLRSLSELEYLMQIPY  
 STTYDIGNFAAFMRKGRDYVHCMPNMDLRDMRHNAAQKDSIELYSLKMAQKILP  
 PYKPCFDKTDQPOSVCSPKFCGCEGSHCTDKVYVALHSLYDI PADEFGAALLR  
 RNHVCAAFHSENLLEDSDYSLDDIGAFPSREGDMLNFSVAESTLNYTHSVN  
 LKVCCTYPASSREVMKEFLVTRVWTFCKESRLDTFVLGVGVHGVGDEKQFVSA  
 MEDAHYKXTLAMNRSERILLEDSSVNTWTFPMKDMVIVPLFDVSLQNEGKGLARKE  
 VMSKDFVYTLNHIHTYQSKALTANVLSFVESRSRVINGVTARSVDWDKALLQ  
 SLSTWFFLQTKLMLKDLVQKFSKSLTEYVWDEITAAAFHNCFTPKERLNNK  
 LITVSEKLEIKVPDLVYTFHDLVKVSKSVEMPVLDVKLSLEABVMYNALESIS1  
 LKDSKEDVDVFSRMCTLGVDPLVAAKVMVAVVSNEGLTTFEFPTEANALALQP  
 TITSKEGSLKIVSSDVGBESSIKEAVRKSEISHLGTLGNTVSDSEFORTEISLQPH  
 MYSTETIRKQHMVYVYVGLKQCKNYLSDLSAASVNLKKI1KDTAIDLDET  
 KEGFYVDYDGLKKLPLSGHGWVNDSDYKCFVALLTYDGENIYVGETWRRVAV  
 SSESIVSDMGKIRAIRSVLKGEPHISAKVTLDVGVPCGCKTKYRLFLDEGLMLHP  
 LVPGQAAEMIRREANSSGLIVATKENVTVDSFLMNYGRGPCQYKRLFLDEGLMLHP  
 GCYNFLVNSLCSCEAFVGTQOIPVINEVATFPVPKHLSOLEVDVAVETRTTLRCPA  
 DITFLNQKTEGQVWCTSSVTSVSHVIOGAAMVNPVSKPLGKVITFTQSDKSLLL  
 SRGVEDVHTVHVQGETFEDVSLVRLTPTPVGLIISKQSPHLLVLSLRHRSIKYITVV  
 LDVAVSVLRDLRCVSVSYLLDMYKVDVSTQ"

## ORIGIN

Query Match 39.8%; Score 2529.6; DB 14; Length 6357;  
 Best Local Similarity 63.7%; Pred. No. 0;  
 Matches 4063; Conservative 0; Mismatches 2259; Indels 58; Gaps 12;  
 QY 6 TTTAATAGTGTTCGACAAACAATATAAACAACAAATATTACAAATATTACAAACAACAA 65  
 Db 6 TTTTTCACATTTAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 64  
 QY 66 CAACAATGGCACACATACATCTATAATTTAGCAACGCCCTTCTTGAAGAGCGTGAGTGA 125  
 Db 65 CTACAATGGCTTACACACAACAAGCTACCAACGCCCATTTAGCAAGTACTCTCCGAGGA 124  
 QY 126 ABAACACTCTCGTTATGACCTTGCAGAGAGCGCATGTACGATACGGCGCTGAGAGAT 185  
 Db 125 ATAAACCTTGGTGAACGACTCTTGTCTAATCGGAGACTGTACGAATCAGCGGTGGAACAAT 184  
 QY 186 TTAACGCCCGCGACCTAGACCAACAGCTCAACTTTTCCAAAACCTATTAGCGAAGAGCAAA 245  
 Db 185 GCATGCAATGACCCGACGCCCAAGCTTAATTTTAAAGTTCGATAAGCGAGAGCAGA 244  
 QY 246 CGCTTCTAGTCTCCACAGCGTACCCGAGGTTCAGATTACCTTTTATATACTCAAAATG 305  
 Db 245 CGCTTATCGCACTAAGGCTACCTCGTGTCCAAATCAGCTTCTACACAACGAGAACG 304  
 QY 306 CGGTACACAGTTGGCTGGAGGTTTGAGAGCATTAGAAATGGGAATATCTGATGCTACAAG 365  
 Db 305 CTGTGCACAGTCTCGCAGGTGGACTTCGGTCTTTTGGAACTAGAAATATTCTGATGATGAGA 364  
 QY 366 TTCCTTATGATCGCGACACATATGATATAGTGGGAACCTTTGAGAGCAGCATTTGTTCAAAG 425  
 Db 365 TCCCTTACGGTTCAACGACATATGATATCGGGGGAATTTTGTCTGCATGATTTTAAAG 424  
 QY 426 GCAGGAGTATCGTGCATTTGCTGTATGCCCAATCTGGACATACGAGATATATATGAGGACG 485  
 Db 425 GTCGTGACTAGCTTTCATTTGCTGATGCCTAACATGACCTTACGTGACGTATCGCTCACA 484  
 QY 486 AAGGACAAAGGAGTCAATTTGATGATGATTTTGTCCAGATGTTCTCGTCTTCTTAACAGGTAA 545







```
Db 4856 TAGTTAAGTACCTGTTGATGATAAAGGTTATTTCAACACATGTTTTCGAGTAAATGGCGT 4915
Qy 4897 TTAGTCTCAGAGATACGTGCAAAATAGCGAGTTCATTTGATCTTTCGAAAACAGGATGAGA 4956
Db 4916 TAGTAGTCAAGACGACGTTAAGATTTCTGAGTTTCATCAATTTGCTGCCCTGAGAAAT 4975
Qy 4957 TACTTCCGCAATTCATGACTAAGTCAAGAGTGTGTAGATATCGACTGTGACACAGATTGA 5016
Db 4976 TCTTACCTGCTGTTATGACTTCCTCGGTCAAGACGTTACGATTTTCGAAAGTTGACAAAGTGA 5035
Qy 5017 TGGCTGTTAAGNATCATAGTCTTTCTGATGTAGATTACTTAAAGGTGTTAAGTTAGTTA 5076
Db 5036 TTGCAATGGAACCAAGATTGTTATCCGATGTGGATTTGCTTAAAGGTGTAAGCTTTGTTA 5095
Qy 5077 AGAAGGAGTGTGTGCTTAGCTGATTTGGTAGTGTCTGGGAGTGGAAATCTCCCGGATA 5136
Db 5096 AGGATGGTATGTGTTTAGCAGGGTTAGTTGTCTCCGGGAGTGGAACTTACCCGACA 5155
Qy 5137 ACTGCCGTGGTGTGCTAGTGTGTATTTAGTAGATAGAGAAATGAAAGGAGTAGGAAG 5196
Db 5156 ACTGACAGAGTGGAGTAAGCGTTTGTGTTGATAGAGAAATGCAAGAGATGACGAAG 5215
Qy 5197 CAACGCTCGGTGCGTATCACGCCCTGCTTCGAAAAGAAATTTTCTTTTAAAGCTAATCC 5256
Db 5216 CAACACTTGGATCTTATAGAACCAAGTGCAGCTAAGAAACGATTTGCGCTTCAAAATTGATCC 5275
Qy 5257 CTAATTAATCAATAACATCCGAGGATGCTGAGAGACCCCGTGGCAAGTGTAGTGAATA 5316
Db 5276 CGAATATAGCATTTACTACCCCGGATGCTGAGAGAAAGTTTGGCAAGTTTGTAGTAAATA 5335
Qy 5317 TCAAGGAGTGGGTATGGAAGAGATCTGCTCTTTATCTTTGAGTGTGTTTCAATTT 5376
Db 5336 TTAGAGTGTGTGCCATGGAAGGTTTCTGCTCTTATCTTTGAGAGTTTGTCTCAGTTT 5395
Qy 5377 GTGTAGTACATAAAATAATGTAAGAAAGTTTGGGACGTAATTTGAGTGTGACAG 5436
Db 5396 GTATTGTACAAATCCATATAAATTAGGCTTGGAGAGAAATTAAGTGTGTCAG 5455
Qy 5437 ACAGCTGCCAATTAATCACTGAAAGGTTGTTGAGGAGTTCGTGGATGAAGTACCAA 5496
Db 5456 AAGGAGGACCGTTGAACTTACAGAGACGTCGTTGATGAGTTCATCGATCAGTTCCAA 5515
Qy 5497 TGGCTGTGAACACTCGAAAGTTCGGAACAAACAAAGAAATGTTAGTGAATAATGTTA 5556
Db 5516 TGGCTGACAGATTACGTAATTTCCCAATCAATCTAAGAAAGGAAGTAATAAGTATGTAG 5575
Qy 5557 ATAATAAGAAATAATAAACAAGTGTGTAAGAGGTTTAAATTTGAGGAAATTCAGGATA 5616
Db 5576 GTAGAGAAATGATAAAGGTTTGAATAAGGAGGAAGCTGTTTGAATAAGGTTAGAA 5635
Qy 5617 ATGTAAGTATGACGAGTCTATCGCGTCATCGAGTACGTTTAAATCAATATGCTTTATAC 5676
Db 5636 TTGGGCAGAACTCGGAGTCATCGAGCCGAGTCTTCTCGTTTAACTATGCTTACAC 5695
Qy 5677 AATCAATCTCCGAGCAATTTGTTTACTTATCTTCGCTTACGAGATCCCTGCGAGCT 5736
Db 5696 AGTTCCAGTGCCCAATCAATTAGTATTTAGTTTCTGTATGGCTGATCCATTAGAGTT 5755
Qy 5737 GATCAATCTGTGACAAATGCAATGGGTAAACCAAGTTTCAACGCAACCAAGCTAGGACAAC 5796
Db 5756 ACAAATCTGTACTTTCGGGTTAGGCATCAGTTTCAACACACACAGCTAGAACTAC 5815
Qy 5797 AGTCCACAGCAATTTCCGAGTCTGGAACCTGTGCTCCTGATGATGACAGTGAATTTCC 5856
Db 5816 GGTTCACACAGCAGTTCTCTGATGTGTGGAAGACTATTTCCAGCCGCTACAGTTAGATTTC 5875
Qy 5857 TGCATCGGATTTCTATGTGTATAGATATAATTCGAGCTGTGCTGCTGATGATGATGATGAT 5916
Db 5876 TGCTACTGGTTTCAAGTTTTCGATATAATGCGGTGCTAGATTTCTTCTAGTGTGCGCACT 5935
Qy 5917 ATTAATAAGCTTTGATACTAGAAATAGAAATAATAGAGGTTTGAATAACAACCCGACCGAA 5976
```

```
Db 5936 TCTCGGACGCTTTGATACTAGGAACAGGATAATAGAGTTGAAAATCCGCAAAATCCTAC 5995
Qy 5977 TACTACTGAATCGTTTAAOCGCACTCAGAGGGTAGAGTACTGTAAGTCTATAAGGGC 6036
Db 5996 AACTGCCGAGACGCTTGATCGACGAGCGGGTAGATGATCGACGCTGCGCATTAGGGC 6055
Qy 6037 TTCAATCAATAAATTTGGCTAATGAATCGTTTCTGGAACCTGCGATGTTCAATCAAGCAGG 6096
Db 6056 CAGTATAGTAACTCATGATGAGTTAGTTCTGGCAGCGGAATGTACAAATCAAGCTCT 6115
Qy 6097 CTTTGAGACTGCTAGTGGACTTGTCTGGACACAACTCCGGCTACTTACGCTATTGTTGTG 6156
Db 6116 GTTCGAGAGCGGAGTGGACTCACCTGGGCTACAACTCCTTAAAC----- 6160
Qy 6157 AGATTTCCTTAAATAAAGTCCGCTGAAGACTTAAATTTCAAGGTTGCTGATCCAAAGTTAAAC 6216
Db 6161 ATGATGGCATAAATAAGTTGGACGAACATTTAAACGCTCC---GTGGCAGTAGCATTAATC 6217
Qy 6217 GCAGTGGTTGCTCGTCCACTTAAATAAAGATTTGTCATATCTGGATCCAAACAGTTAAAC 6276
Db 6218 GTAGTGTTTTCCCTCCACTTAATCGAGGGTTGCTGTTGGATGGAACGCAATTAAT 6277
Qy 6277 CATGTGATGTGTATCTGTGATGCGGTAA--AACATCGAGAGGTTTCAATCTCTCCC 6335
Db 6278 ACATGTGTGACGTGTATTTTCCGAAACGACGTAATTTTTCAGGGGTTCCGAATCCCCC 6337
Qy 6336 TAACCCGCGGTAGCGGCCCA 6355
Db 6338 GAACCCGCGGTAGCGGCCCA 6357
```

## RESULT 15

```
PPPSLF LOCUS DEFINITION
Nucleotide sequence of the genomic RNA of the tobamovirus Pepper
M81413.1 GI:333293
coat protein; replication-associated protein.
Pepper mild mottle virus
Pepper mild mottle virus
VIRUS; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
1 (bases 1 to 6357)
Alonso, E., Garcia-Luque, I., de la Cruz, A., Wicke, B.,
Avila-Rincon, M.J., Serra, M.T., Castresana, C. and Diaz-Ruiz, J.R.
Nucleotide sequence of the genomic RNA of pepper mild mottle virus,
a resistance-breaking tobamovirus in pepper
J. Gen. Virol. 72 (Pt 12), 2875-2884 (1991)
92113528
1765765
Original source text: Pepper mild mottle virus (strain S) cDNA to
genomic RNA.
Location/Qualifiers
1..6357
/organism="Pepper mild mottle virus"
/mol_type="genomic RNA"
/strain="S"
/db_xref="taxon:12239"
70..4908
/note="183K ORF; internal stop codon"
/codon_start=1
/transl_except=(pos:3421..3423,aa:OTHER)
/product="replication-associated protein"
/protein_id="AAB02334.1"
/db_xref="GI:333295"
/translacion="WAYTQATNAALASTLRGNPLVNDLANRLIYSAVQCNAHDR
RPKNPLRSISEQTLLATKATPEFQITFTNTQAVSLAGLSLELEYLMMQIPY
STTYDIGNFAHRPKGDIYVHCMMDLRVMRHAQKDSIELYSLKAKKKVIP
PYQKCFDLYDDPSVVCSPKQFQCEGSHCTDKYVALHSLYDIPADFGAALLR
RNHVCAAFHSENLILDSYSLDDIGAFSREGDMLNFSVAESTLNTHTSYNV
LKVCYKTFPASREVVKBFVTRVNTWFKFSRLDTFVLYRGVTHRGVDEKQFSA
MEDAHYKKTLLAMNSRILLESSSVNYNPFPMKDMVIVLPFVLSIQNEKRLARKE
VWVKDFVYTVLNHRTQSKALTYANVLSFVESIRSKVITINGVTARSEVDVDAKLLQ
```



```
LSWSTPFLQTKLAWLKDVLVQKFOVHSKSLTEYVWDELTAAPHNCPTTKERLNKK
LITVSEKALEIKVPDLTYTFHDLRLVKEYKSSVEMPVLVDKKSLEAEAEVWMLSEISI
LKDSKDFVDVFERMCNTLIGVPLVAAKVWVAVVSNESGLTLTFEPTBANVALALQP
TITSKEGSLKIVSSDVGSSIKVRYKSEISMLGTLGNTVSDSFORSFTEISLQOPH
MVSTETIRKQKHAMVYTGPKLVQCKNYLDSLVASLSAAVSNLKKIIKDTAIDLDET
KEKFGVYDCLKKWLKPLSKGHAWGVWMSDYKFCFVALLTYDGENIVCGETWRRVAV
SSSLVYSDMGKIRATRSVLKDGEPHISAKVTLVDVPGCGKTKELIRVNFDEDLV
LVPKQAAEMIRRRANSGLIVATKENVRITVDSPLMNYGRGPCQYKRLFDGLMLHP
GCNFIIVGMSLSEAFYIGDTQIPIYNIRVATFPYKHLSQLSLEQVDAVETRRITLRCFA
DITFFLNQYKEGVMCTSSVTSVSHVEIQAAMVNPVSKPLGKVITFTQSDKSLLL
SRGEYDHTVHYEQGTFFEDVSLVRLTPTPVGIISQSPHLLVSLSRHRSRIKYITV
LDAVSVLRDLECVSSYLLDMYKVDVSTQXQLOIESVYKGVNLFVAAPKTDGVSDMOY
YDKCUPGNSITLINEYDAVMTQIRENSLVKDCVLDMSKSVLPRESEITLKEPVIRTA
AKERPKGLENLVAMKKNFNSPELVGVVDIEDTASLVVDKPEFDAYLKEKKKPKNI
PLLSRASLERWIKSEKSTIGQADFDIDLPAVQQRMIKQPKQRDLDSLTQITP
ALQTIIVHSKINKALGPFVSELTROYELLETIDSSRFMYTRKPTQITBEFFSDLSNV
PMDILBLSKYDKSQNEHFCAVEYBIWRKLDGLDDFLAEVWKHGRKTLTKDYTAGIK
TCLWYQRKSGDVTFITGNTIIIAACSLMPEMLIKGAFCDGDDSLPYPKGTDPDI
OQGANILMNPFAKLPKRYGYPCGRVI IHHDRGCI VYVDPLKLI SKLGAKH IKNRHL
EEFRTSLCDVAGSLNANCAYTHLNDVAGEVITAPLGSFVYRALVLYLCKRLPQLFL
LE"
70..3423
CDS
/notes="126K ORF"
/codon_start=1
/product="replication-associated protein"
/protein_id="AAB02335.1"
/db_xref="GI:333294"
/translation="MAYTQATNAALASTLRGNPNPLVNDLANRRLYESAVQCNADR
RPKNVLRGISSEQTIIATKAYPEFQITFYNTQNAVHSLAGLSLELSYLMWQIPIYG
STYDIGNFAAPMQRGVVHCCPMNDLRDVMRHNAOKDSIELVLSKLAOKKVIPI
PVQKPCFDKYDQPSVCSKPKQCEGYSHCCTKYVAVALHSLVDIPADEGAALLR
RNHVCTIAAFHSENLLSDSYSLDIDGAPFREGDMLNPFVAESTLNTSHSYSNV
LKVKCTYTPASSRSEYKELFVTRVNTWPKFSLDITFVLYGVGVHGRGVKQEFYSA
MEDAHYKXTLAMNREILLEDSSVNTYFPRMKDMVIVPLFDVSLONEGRKLARKE
VMSKDFVYTLNHIHTYQSALTYANLSFVBSIRSRVINGVITARSBDVVKALLO
LITVSEKALEIKVPDLTYTFHDLRLVKEYKSSVEMPVLVDKKSLEAEAEVWMLSEISI
LKDSKDFVDVFERMCNTLIGVPLVAAKVWVAVVSNESGLTLTFEPTBANVALALQP
TITSKEGSLKIVSSDVGSSIKVRYKSEISMLGTLGNTVSDSFORSFTEISLQOPH
MVSTETIRKQKHAMVYTGPKLVQCKNYLDSLVASLSAAVSNLKKIIKDTAIDLDET
KEKFGVYDCLKKWLKPLSKGHAWGVWMSDYKFCFVALLTYDGENIVCGETWRRVAV
SSSLVYSDMGKIRATRSVLKDGEPHISAKVTLVDVPGCGKTKELIRVNFDEDLV
LVPKQAAEMIRRRANSGLIVATKENVRITVDSPLMNYGRGPCQYKRLFDGLMLHP
GCNFIIVGMSLSEAFYIGDTQIPIYNIRVATFPYKHLSQLSLEQVDAVETRRITLRCFA
DITFFLNQYKEGVMCTSSVTSVSHVEIQAAMVNPVSKPLGKVITFTQSDKSLLL
SRGEYDHTVHYEQGTFFEDVSLVRLTPTPVGIISQSPHLLVSLSRHRSRIKYITV
LDAVSVLRDLECVSSYLLDMYKVDVSTQ"
4909..5682
/misc_feature="cell to cell spreading of the virus"
/notes="30K ORF"
/codon_start=1
/protein_id="AAB02336.1"
/db_xref="GI:333296"
/translation="MALVVKDDVKI SEFINLSAERKFLPAVMTSVKTVIRISKVDKVI
A MENDSLSDVNLKGVLKVDGYCLAGLVSGENWLPDNCRGVSVCLVDKQNRDDE
ATUGSTRISAARKFAPKLPNTSITADAEKRVQVQVLRVIRGVAMEKGFCLPSLEFV
SVCIVHKNKILGRLKRTISVSEGGVPELVEAVVDEFIENVPMADRLRFRNQSKKGS
NKTVGRNDNKGRLKEGLFDKVRIGQNSSESDAESSE"
5458..5517
misc_feature
/notes="assembly origin"
5685..6158
/misc_feature="coat protein"
/product="coat protein"
/protein_id="AAB02337.1"
/db_xref="GI:333297"
/translation="MAYTVSSANQLVYLGSWADPLELQNLCTSALGNQFOQOQARTT
VQOQFSDVWKTIPTATVPATGKVFYRVNAVLDLSVALGAFDRNRRIIEVENPON
PTTAETLDATRDVDAIVAIRASISNLMLNELVRGTGMYNQALPESASGLTWATP"
6159..6357
misc_feature
/notes="3'-end noncoding region (tRNA-like sequence)"
```

```
ORIGIN
Query Match 39.8%; Score 2528.4; DB 14; Length 6357;
Best Local Similarity 63.6%; Pred. No. 0;
Matches 4056; Conservative 0; Mismatches 2266; Indels 58; Gaps 11;

QY 6 TTTAATAGTTTTCGACAACAACAAATTTAAACAAAACAAACATATTACAAACAAACAA 65
DB 6 TTTTTCACAAATTTAAACAACAACAACAACAACAACAACAACAACAACAACAACA 64
QY 66 CAAACATATGCGACACACATACATCTATAATTTAGCAACACCCCTTCTTTGAAAGCGTGAGTGA 125
DB 65 CTACAAATGCTTACACACACAACAAGCTACCAACGCGGCATTTAGCAAGTACTCTCCGAGGA 124
QY 126 AAAAACTCTCGTTAATGACCTTTCAGAAAGGCGCATGTACGATCGGCGCGTGAAGAAT 185
DB 125 ATAACCCCTTGGTGAACGATCTTGTCTTAAATCGGAGACTGTACGAATCAGCGGTGCAACAAT 184
QY 186 TTAAGCCCGCGACCGTAGACCAACAAAGGTCAACTTTTCCAAAACATATTAGCGAAGAGCAAA 245
DB 185 GCATGCAATGACCCGAGGCCAAGGTTAATTTTTTAAGTTCGATAGCGAAGAGCAGA 244
QY 246 CGCTTCTAGTCTCCAAACGCGTACCCGAGGTTCAGAGATTACCTTTTATAATCTCAAAATG 305
DB 245 CGCTTATCGCAACTAAGGCTTACCCCTGAGTTCCAAATCAGTTCTTACAACACGCGACAACG 304
QY 306 CGGTACACAGTTTGGCTGGAGGTTTTCAGAGCATTTAGAAATTTGGAAATCTTGATGCTCAAG 365
DB 305 CTGTGACAGTCTCGCAGGTGGACTTCGGTCTTTTGGAACTAGAAATCTTGATGATGACGA 364
QY 366 TTTCCCTATGATCGCGACATATGATATAGTGGGAACCTTTGCAGCACATTTGCTTCAAG 425
DB 365 TCCCTTACCGTTCAACGACATATGATATCGGGGGAATTTTGTGCTCACATGTTTAAAG 424
QY 426 GCAGGATTACGTGCACTTCTGTATGCCCAATCTCGACATACAGAGATATAATAGGACCG 485
DB 425 GTCTGACTACGTTCACTGCTGATGCTTAAATGCACTTACGTGAGTCACTGCGTCA 484
QY 486 AAGGCAAAAGGACTCAATTTGATGATATTTGTCCAGATTTGCTCTCTTCTTAAACAGGTAA 545
DB 485 ATGCTCAAAAGGATAGCAATTTGAATCTTCAAAAGCTTGGCGAAAAGAAAAGGTAA 544
QY 546 TTTCTGAGTTTCAAGGGAGGCTTTTAAACAGGTATGTCAGAGCTCCCAACGAGTCTGCT 605
DB 545 TACCGGCATATCAAAAGCCATGCTTTGATAAATACACGACGATCCGCAATCACTAGTAGT 604
QY 606 GCTCTAAAACTTTTCAAGGATTTGCGAATATACATCCCGCAGAGATAGTGGTAGAAGATACG 665
DB 605 GCTCGAAACCTTTTCAAGCACTGCGAAGGCTTTGCG--ACTGCACGGATAAAGTATACG 661
QY 666 CTGTTGCTCTGCACAGTTTGTATGATATTTCTGTGTGATGAGTTTGGAGCTGCGTTAATAT 725
DB 662 CTGTGCTTTGCACAGTTTATACGACATTTCCAGCAGATGAAATTTGGGGCAGCACTTCTGA 721
QY 726 CTAAGAAATATACATGATGTTATGAGCTTCCATTTTGGCAGAGCATTTATTACTAGACC 785
DB 722 GGAGAAATGTTCAATGCTCTGATGCTGCTTCCACTTTTCTGAGAAATCTTCTTTTGAAG 781
QY 786 AGACGGAGGTTACGCTTAATGAAATAGGCGCAACTTTTCAAAAAGAGAGGTGATGATGTTT 845
DB 782 ATTCGTATGTCAGTCTTGACGACATAGGCGCTTTCTTCTCGAGAGAGGGCGGATATGTA 841
QY 846 CTTTTTTCTTTGCTGATGAAGTACTTTTAAATATATAGTCAATAAATACAAAATATCTTGC 905
DB 842 ACTTTTCTTTTGTAGCAGAGAGTACTTTTAAATATATACTCATCTTATAGTAAATGTGCTTA 901
QY 906 ATATGATAGTAAATCTTACTTTCTGCTTCTAGTAGAATAGTTTATTAAAGAAATTTT 965
DB 902 AGTATGTTGTGAAGTACTTCTCCCGCTTCTTAGTAGAAGGTGATCATGAAGAGGTTTT 961
QY 966 TAGTCTACTAGGGTTAATACTTGGTTTGTAAATTTTACCAGAAAGTAGATACCTATATCTGT 1025
DB 962 TGGTAACTAGGGTAAATACTTGGTTTGTAAAGTTTCAAGGTTAGATACCTTTGTACTAT 1021
```

1026 ACAAGAGTGTAGACAAGTAGGGTGTGATAGTACAGTCTTATGAGCGGATGGAAGACG 1085  
1022 ATAGAGGTGTATACCAACAGAGGTGTAGACAAGGAGCAATTTTACAGTCAATGGAAGATG 1081  
1086 CTTTGGCTTACAAAGAAACCTTGGCCATGTTTCAACACCTGAAAGAGCAATCTTTAGAGACA 1145  
1082 CTTGGCATTACAAAAGAGCTTTGGCGATGATGATAGCGAAAGAAATCCTCTTAGAGGATT 1141  
1146 CGGCTTCGGTTAACTTTTGGTTCCCTAAGATGAAGGACATGGTGATAGTACCGCTGTTTG 1205  
1142 CATCGCTCTGTTAATTATTGGTTTCAAAGATGAAGATATGGTGATAGTACCTTTGTTTCG 1201  
1206 AGGGTTCTATTACCAAGC-----AAAAAGATGACAAGGAGTGAGTCAATTTGTTAATCGTG 1259  
1202 ACGTATCTTTACAGAACGAGGGAAGAGTTAGCAAGAAAGAGGCTCATGTCAGCAAGG 1261  
1260 ACTTCGTTTACACAGTGTCTTAATCATATCAGNACATATCAAGCCAAAGCGTTAACTTACC 1319  
1262 ACTTCGTTTACTGTGCTTAATCATATTCGCACATACCACTCGAAAGCGCTTACTTACG 1321  
1320 AGAAGTATTATCTTTCCGTGGAGTCTATAAGATCCCGGTGATTAATCAATGGTGTACTG 1379  
1322 CCAATGTTATATCGTTCTGAGTCGATAAGATCAAGAGTGATTAATCAATGGGAGTACTG 1381  
1380 CTAGGTCTGAATGGGATGTATGAATAAGCAATTTCTTCAACCTTGTCAATGACTTTCTTCT 1439  
1382 CGCGCTCAGAGTGGGATGTGATAAGGCTTTGTTGCAAGTCCCTGTCAATGACTTTTTTCT 1441  
1440 TGCAGACTAAGCTGCTCGCTTCAAGACGATATAGTAATGGGAAGTTTCGGTGTCTGG 1499  
1442 TGCAGACCAATTTGGCCATGCTCAAGGATGACCTCGTGTGTTTCAGAAATTCAGTGCATT 1501  
1500 ATAAGACCACTTCTGAACCTATTATTTGGGATGAGGTGGGCAAAATTTTGGAAAGCTTTTCC 1559  
1502 CCAATCGCTCACTGAATATGCTGGGATGAGATTACTGTGCTTTTTCACAAATTTGTTTC 1561  
1560 CCACTATCAAGAGAGATTGTGTAGCAGGAAATTTCTGGATGTAAGTGAGAACTGCTCTGA 1619  
1562 CTACAATCAAGGAGAGGTGTGATTAACAAGAACTCTATACTGTTTTCGGAAGAGCTCTTG 1621  
1620 AGATCAAGATCCAGATCTGTATGTCATGGAAGACAGTTCGTAGCTGAAATACACCA 1679  
1622 AAATTAAGTACCTGATTTGTATGTAATCTTCCAGATAGATTGTTTGAAGAGTACAAAT 1681  
1680 AGTCTGAGGAGTTACCGCATCTAGATATCAAGAAAGCTTTAGAAAGAGCTGAGCAATGT 1739  
1682 CTTCCGTTGGAATGCCGTAAGTCTGACGCTTAAAGAGAGCTTGAAGAAAGAGAGTGTAT 1741  
1740 ACGAGCGGTTATCAGAAATATCTATCTTAAGGTGCTGATTAATTTCCGATATCGGAGT 1799  
1742 ACAATGCTTTGTCAAGAACTCTCAATCTTAAAGACAGTGAACAAGTTTGTATGTTGATGTT 1801  
1800 TCAAAGACATGTGCAAGGCTTTAGATGTTAGTCTGATGTGGCAGCACAGGTAATCGTTG 1859  
1802 TTTCCCGGATGTGTAATACATTAGGCGTAGATCCATTGGTGGCAGCAAGGTAATGTAG 1861  
1860 CAGTGGCCGAGATAGAAAGCGGTTTAACTCTTACTTTTGATAAGCCACCGAGGAGAAATG 1919  
1862 CTGTGGTTTCAAATGAGAGTGGTTTGACCTTAACTTTGAGAGGCTTACCAGAGCAATG 1921  
1920 TGGCTAAGGCTCTTAAAGCAGCGGCTGAGGCGGTGGTATGT-----CTTGAAC 1970  
1922 TCGCACTTGCATTGCCAACCGCAATTTACATCAAGGAGGAGGTTGCGTTGAAGATTGTGT 1981  
1971 CGCATCCGAGAGGTGAACGTAAATAAATTTTCTATTGCTGAGAAGGAGATTCGCTG 2030  
1982 CGTCAGACGTAGGTGAGTCTCTCAATCAAGGAAGTGGTTCCGAATACAGAGATTCTATGC 2041  
2031 TGTGTGAGAAAGTCTAGTTTGAAGATGCTAACTTAGAGCACCGAGGAGTTGAGTCCC 2090  
2042 TTGGTCTAAAGGCAACACAGTGTCCGATGAGTTTCCAAAGAAAGTACAGAAATCGAGTCTG 2101

2091 TCACAGATTTCCATAAGGCTTGGTGGATAGTGTGATTACAAAGCAAAATGGCATCGTTG 2150  
2102 TGCAGCAGTTCCATATGTTATCCACAGAGACGATTATCCGTAACAGATGTCATGCCATGG 2161  
2151 TCTACACTGGCTCACTCAAAGTTCAACAAATGAAGAACTATGTGGAAGTTTGGCAGCTT 2210  
2162 TGTATACTGTGTCGGCTAAAAGTTCAACAAATGCAAGAACTATTTAGACAGCGCTGGTAGCCT 2221  
2211 CGTTGTCCGCACTGTATCAAAATCTATGCAAGTCACTAAAGGATGAAGTCGGGTATGATT 2270  
2222 CGTCTCTGTGTCGGTATCAACCTGAAGAAGATAATCAAGACACAGCTGTATAGATC 2281  
2271 CTGATTTCCAGGAGAAAGTTGGTGTGGATGTCACTTTTGAAGAAAGTGGCTCTCAAAAC 2330  
2282 TCGAGACTAAGGAAATTTGGAGTCTACGAGTGTGCCCTTAAGAATGGTTGGTGAAC 2341  
2331 CTGCGGCCAAAGTCAATCAATGGGAGTTGTCTCTGATTAACAAGGGGAAATATGTTTACTG 2390  
2342 CTCTATCAAAAGGACATGCTTGGGTGTGGTGTATGAGCTCAGACTATAAGTGTCTTGTG 2401  
2391 CACTTCTATTTATGAAGGAGATAGAAATGGTGTGACTGAGAGCGACTGGAGAGGGTGGCTG 2450  
2402 CGCTTCTACATACGATGGCGAGAACATTTGTGTGGGAGAGACATGGCGTAGTGCAG 2461  
2451 TATCATCTGATACAATGGTATATTTCTGATATTGCAAAAGCTCCAAAATCTGAGGAAACAA 2510  
2462 TGAGCTCCGAATCTTTGGTGTATTCAGATATGGGAGATAGAGCTATACGCTCTGTGC 2521  
2511 TGGAGACCGGTGACCCCAAGAACCTACTGTGCAAGATGTGTACTTGTGGATGGGTGGCTG 2570  
2522 TTAAGACCGGTGAACCCCATATAAGCAGTGCAGAGGTTACACTTGTGTGATGGTGTCTG 2581  
2571 GTTGTGGAAGTACAAAGGAGATTTTGAAGATTTGATCTTCATGAGGATTTGATCTTGG 2630  
2582 GTTTCGGAAGACAAAGAGATCTTTTCAGGGTCAACTTTGACGAGATCTAGTCTTGG 2641  
2631 TTTCTGGAACCAAGCTGTCTGTATGATCAGAAAGAGGCTAAATTCATCTGAGCTGATAA 2690  
2642 TACCAGGAAACAGGCTGTGAAATGATAAGAAAGAGGCAACAGTTCTGGTTTAAATCG 2701  
2691 GAGCCACATGACATGTGAGAACGGTAGATTCACTTCTATGCA-----TCCAAAC 2744  
2702 TGGGACCAAGGAGATGTAGAGCGTAGACTCTTCTTATGAAATTAACGCTCAGGTC 2761  
2745 CGCGATCACACAAGAGGCTTTTATTTGATGAAGGTTGATGCTGCACACCGGTTCTGTTA 2804  
2762 CGTGCCTAACAAAGGCTTTTCTGGATGAAGTCTAATGTTACACCTGTGTTGTTA 2821  
2805 ACTTCTGTGCTTATCTCTGTGTCGACATCGCATACATTTACGGAGATACACAGCAGA 2864  
2822 ATTTTCTGGTGGCATGCTCTATGCTCCGAGGCTTTTGTGTTATGAGACACCCAGCAGA 2881  
2865 TTCTCTTCAATTAACAGATTTCAGAAATTTCCGATATCCCAACATTTTGAAGCTGCAAG 2924  
2882 TTCTTTACATCAACAGATTGCAACTTTTCCCTATCTTAAAGCATTTGAGTCAACTCGAGG 2941  
2925 TGGATGAAGTTTCAGATGAGGAGGACCACTGAGATGCCAGGTTGATGTAATTTTTC 2984  
2942 TCGATGCTTTGAACCTCGCAGAACAGTTCGGGTGTCAGCTGATATCACCTTCTTCT 3001  
2985 TACAATCGAAGTACGAAGGAGCGGTGACAAACATTCATCTGTGTACAGCATCGGTCTCAT 3044  
3002 TGAATCAGAAGTACGAAGGCGCAAGTTATGTGCATCAAGTGTGTACACGCTCGGTGTAC 3061  
3045 CTGAGATGATAGGCGGTAAAGGAGTAAACAGTGTTCCAAACCACTTAAAGGGAAGA 3104  
3062 ACGAGTCTCAAGAGTGCAGCGGTATGATTCAGTGTCTTAACCTTTAAGGGAAGG 3121  
3105 TTGTAACCTTCACTCAGGCTGATAAATTTGAGTTAGAGGAGAGGCTTATAGAAATGTGA 3164  
3122 TGAATACATTCAGTCAAGTCAAGTCAATGCTCTCGAGGGGTACCAAGATGTGC 3181  
3165 ACACCGTTTCATGAGATCCAAGGAGAAACCTTTGAAGATGTGTGCTGCTGAGATTGACGG 3224





GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 16, 2005, 21:45:12 ; Search time 1904.23 Seconds  
(without alignments)  
17518.955 Million cell updates/sec

Title: US-09-551-494-5  
Perfect score: 6355  
Sequence: 1 gatgtttaatagtttcga.....taaccgcgtagcgccca 6355

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_23Sep04:\*

- 1: Geneseq1980s:\*
- 2: Geneseq1990s:\*
- 3: Geneseq2000s:\*
- 4: Geneseq2001as:\*
- 5: Geneseq2001bs:\*
- 6: Geneseq2002as:\*
- 7: Geneseq2002bs:\*
- 8: Geneseq2003as:\*
- 9: Geneseq2003bs:\*
- 10: Geneseq2003cs:\*
- 11: Geneseq2003ds:\*
- 12: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6355	100.0	6355	3	AAC62372 CDNA sequ
2	6355	100.0	6355	9	Acc85005 TMV-U2 ge
3	2631	41.4	6384	2	Aaq71546 Component
4	2621.6	41.3	6384	6	AbA04211 Tomato mo
5	2468	38.8	6395	2	Aaz20642 TMV-based
6	2468	38.8	6395	4	Aaf82330 Wild-type
7	2466.4	38.8	6395	2	Aaq95155 Tobacco m
8	2466.4	38.8	6395	3	AAC62369 CDNA sequ
9	2455.4	38.6	6439	2	Aaz20643 TMV-based
10	2455.4	38.6	6439	4	Aaf82331 Tobacco m
11	2452.2	38.6	6475	2	Aaz20645 TMV-based
12	2452.2	38.6	6475	4	Aaf82333 Tobacco m
13	2449.8	38.5	6446	2	Aaz20646 TMV-based
14	2449.8	38.5	6446	4	Aaf82334 Tobacco m
15	2431.2	38.3	6425	2	Aaz20644 TMV-based
16	2431.2	38.3	6425	4	Aaf82332 Tobacco m
17	2291.4	36.1	11222	12	Adn97500 Artificial
18	2291.4	36.1	11641	6	Aad45229 TTODA (rg
19	2291.4	36.1	11641	6	Aad30539 TTODA rGA
20	2291.4	36.1	11641	10	Add84772 Tobacco m
21	2291.4	36.1	11641	12	Adm48702 TMV based

22	2289.8	36.0	10600	11	ADM68432	Adm68432 Celery CE
23	2289.8	36.0	10600	12	ADP26585	Adp26585 TMV CEL I
24	2289.8	36.0	10624	11	ADM68433	Adm68433 Celery CE
25	2289.8	36.0	10624	12	ADP26586	Adp26586 TMV CEL I
26	2289.8	36.0	11641	12	ADJ88298	Adj88298 TTODA chi
27	2288.2	36.0	7685	5	AAD20298	Aad20298 Tobacco m
28	2288.2	36.0	7685	5	AAD20294	Aad20294 Tobacco m
29	2288.2	36.0	7685	6	AAD24478	Aad24478 Tobacco m
30	2288.2	36.0	7685	6	AAD24482	Aad24482 Tobacco m
31	2286.6	36.0	7685	4	AAD02010	Aad02010 TMV viral
32	2286.6	36.0	7685	5	AAD20288	Aad20288 Tobacco m
33	2286.6	36.0	7685	5	AAD20291	Aad20291 Tobacco m
34	2286.6	36.0	7685	6	AAD24472	Aad24472 Tobacco m
35	2286.6	36.0	7685	6	AAD24475	Aad24475 Tobacco m
36	2286.6	36.0	7685	10	ADB83319	Adb83319 DNA trans
37	2286.6	36.0	7685	10	ADB81243	Adb81243 DNA trans
38	2286.6	36.0	7685	10	ADB81246	Adb81246 DNA trans
39	2286.6	36.0	7686	5	AAD20289	Aad20289 Tobacco m
40	2286.6	36.0	7686	6	AAD24473	Aad24473 Tobacco m
41	2286.6	36.0	7686	8	ADA15012	Ada15012 Tobacco m
42	2286.6	36.0	7686	10	ADB83320	Adb83320 DNA trans
43	2286.6	36.0	7686	10	ADB81244	Adb81244 DNA trans
44	2286.6	36.0	7687	5	AAD20290	Aad20290 Tobacco m
45	2286.6	36.0	7687	6	AAD24474	Aad24474 Tobacco m

## ALIGNMENTS

## RESULT 1

AAC62372  
ID AAC62372 standard; DNA; 6355 BP.

XX AC AAC62372;

DT 19-MAR-2001 (first entry)

XX cDNA sequence of the genome of tobacco mosaic virus-U2.

DE Plant phenotype; gene trait; Nicotiana; Oryza sativa; Zea mays; Brassica;  
KW Gossypium; Triticum; Arabidopsis; Petunia; herbicide; transgenic plant;  
KW tobacco necrosis virus; TMV; tobacco mosaic virus; TMV; helper virus; ss.

OS Tobacco mosaic virus.

XX WO200063397-A2.

XX PD 26-OCT-2000.

XX PF 17-APR-2000; 2000WO-EP003521.

XX PR 20-APR-1999; 99US-00294022.

XX PA (AVET ) AVENTIS CROPS SCIENCE NV.

XX PI Meulwaeter F, Cornelissen M, Jacobs J, Van Eldik G, Metzlaiff M;

XX DR WPI; 2000-687182/67.

XX PT Identifying and isolating genes involved in determining the trait or  
PT phenotype of plant species, by infecting plants with gene silencing  
PT constructs targeted to the gene, and identifying plants with altered  
traits.

XX PS Example 1; Page 53-56; 64pp; English.

XX CC The specification describes a method for isolating genes that determine a  
CC trait or phenotype of a plant species. The method comprises identifying a  
CC set of nucleic acids of genes correlated with the trait, creating a  
CC library of gene silencing constructs in a viral RNA vector, targeting the  
CC gene silencing constructs to the nucleic acid set, infecting a collection  
CC of individual plants with these, identifying plants with altered traits  
CC or phenotype, and isolating genes of the invention. The method is useful



Db 1861 AGTGGCCGAGAAATAGAAAGCGGTTTAACTTTATCTTTTGTATAGCCAAACCGAGGAGAAATGT 1920  
Qy 1921 GGCTAAGGCTCTTAAAGACAGCGGCTCTGAGCGCGTGTATGTCTTGAACCGACATCCGA 1980  
Db 1921 GGCTAAGGCTCTTAAAGACAGCGGCTCTGAGCGCGTGTATGTCTTGAACCGACATCCGA 1980  
Qy 1981 AGAGGTGAACGTAATAATTAATTTTCTATTTCTGAGAAAGGAGATGGCTGTGTGTGAGAA 2040  
Db 1981 AGAGGTGAACGTAATAATTAATTTTCTATTTCTGAGAAAGGAGATGGCTGTGTGTGAGAA 2040  
Qy 2041 AAGTCATGGTTTGAAGAACTGTAATTTAGACACACAGAGTTGGAGTCCCTCAACGATTT 2100  
Db 2041 AAGTCATGGTTTGAAGAACTGTAATTTAGACACACAGAGTTGGAGTCCCTCAACGATTT 2100  
Qy 2101 CCATAAGGCTTGGCTGGATAGTGTATTAACAAAGCAAAATGGCATCGGTGTGTCTACACTGG 2160  
Db 2101 CCATAAGGCTTGGCTGGATAGTGTATTAACAAAGCAAAATGGCATCGGTGTGTCTACACTGG 2160  
Qy 2161 CTCACTCAAGTTCAACAAATGAAGAACTATGTGTGACAGATTTGGCAGCTTGGTGTCCGC 2220  
Db 2161 CTCACTCAAGTTCAACAAATGAAGAACTATGTGTGACAGATTTGGCAGCTTGGTGTCCGC 2220  
Qy 2221 CACTGTATCAAACTATGCAAGTCACTAAAGGATGAAGTGGGTATGATTTCTGATTTCCAG 2280  
Db 2221 CACTGTATCAAACTATGCAAGTCACTAAAGGATGAAGTGGGTATGATTTCTGATTTCCAG 2280  
Qy 2281 GGAGAAAGTTGGTGTGGGATGTCACTTTGAAAGAGTGGCTCTCAAACTCGGGCCAA 2340  
Db 2281 GGAGAAAGTTGGTGTGGGATGTCACTTTGAAAGAGTGGCTCTCAAACTCGGGCCAA 2340  
Qy 2341 AGGTCAATTCATGGGGAGTTGCTGGATTAACAGGGGAAATGTTTACTGCACTTCTATC 2400  
Db 2341 AGGTCAATTCATGGGGAGTTGCTGGATTAACAGGGGAAATGTTTACTGCACTTCTATC 2400  
Qy 2401 TTATGAAGGATAGAAATGGTGAATGTGAGAGGAGTGGGTGATCATCTGA 2460  
Db 2401 TTATGAAGGATAGAAATGGTGAATGTGAGAGGAGTGGGTGATCATCTGA 2460  
Qy 2461 TACAATGGTATATCTGATATTCGAAGCTCCAAATCTGAGGAAACAAATGAGAGACGG 2520  
Db 2461 TACAATGGTATATCTGATATTCGAAGCTCCAAATCTGAGGAAACAAATGAGAGACGG 2520  
Qy 2521 TGAACCCACGAACTTACGAAAGATGTTGTTGATGGGTGGCTGTGTGGAAA 2580  
Db 2521 TGAACCCACGAACTTACGAAAGATGTTGTTGATGGGTGGCTGTGTGGAAA 2580  
Qy 2581 GTACAAAGGATATTTGAAGATTTGATCTTGTGATGAGGATTTGATCTTGGTTCCTGGAAA 2640  
Db 2581 GTACAAAGGATATTTGAAGATTTGATCTTGTGATGAGGATTTGATCTTGGTTCCTGGAAA 2640  
Qy 2641 ACAAGCTGCTCTATGATCAGAAGAGGGCTTAATTCATCTGGAATGATTAAGAGCCAAAT 2700  
Db 2641 ACAAGCTGCTCTATGATCAGAAGAGGGCTTAATTCATCTGGAATGATTAAGAGCCAAAT 2700  
Qy 2701 GGAACAATGTGAGAACGGTAGATTCATCTTAATGCAATCCAAAACCGCGATCACAAAGAG 2760  
Db 2701 GGAACAATGTGAGAACGGTAGATTCATCTTAATGCAATCCAAAACCGCGATCACAAAGAG 2760  
Qy 2761 GCTTTTATGATGAAGGGTGTGATGCTGCAACCGGTTGTGTTAACTTCCCTGGTGTAT 2820  
Db 2761 GCTTTTATGATGAAGGGTGTGATGCTGCAACCGGTTGTGTTAACTTCCCTGGTGTAT 2820  
Qy 2821 CTCTGGTTGCAACATCGCATATTTACGGAGATACACAGCAGATTCCTTTTCAATTAACAG 2880  
Db 2821 CTCTGGTTGCAACATCGCATATTTACGGAGATACACAGCAGATTCCTTTTCAATTAACAG 2880  
Qy 2881 AGTTCAGAAATTTCCGATATCCCAACATTTTGAAGCTGCAAGTGGATGAAGTTGAGAT 2940  
Db 2881 AGTTCAGAAATTTCCGATATCCCAACATTTTGAAGCTGCAAGTGGATGAAGTTGAGAT 2940  
Qy 2941 GAGGAGGACCACTGAGATCCCAAGTGTGATTTTCTTACAAATCGAAGTACGA 3000  
Db 2941 GAGGAGGACCACTGAGATCCCAAGTGTGATTTTCTTACAAATCGAAGTACGA 3000

Qy 3001 AGGAGCGGTGACAAACCACTTCAACTGTATCAACGATCGGTCTCATCTGAGATGATAGCGG 3060  
Db 3001 AGGAGCGGTGACAAACCACTTCAACTGTATCAACGATCGGTCTCATCTGAGATGATAGCGG 3060  
Qy 3061 TAAGGGAGTACTAAACAGTGTCTTCCAAACCACTTAAAGGGGAAATTTGTAACCTTCACTCA 3120  
Db 3061 TAAGGGAGTACTAAACAGTGTCTTCCAAACCACTTAAAGGGGAAATTTGTAACCTTCACTCA 3120  
Qy 3121 GGCTGATAAATTTTGAGTTAGAGGAGAAAGGGCTATTAAGAATGTGAAACCGGTTCAATGAT 3180  
Db 3121 GGCTGATAAATTTTGAGTTAGAGGAGAAAGGGCTATTAAGAATGTGAAACCGGTTCAATGAT 3180  
Qy 3181 CCAAGGAGAAACCTTTTGAAGATGTGCTGCTGAGATTTGAACGGCAACTCCACTGACTCT 3240  
Db 3181 CCAAGGAGAAACCTTTTGAAGATGTGCTGCTGAGATTTGAACGGCAACTCCACTGACTCT 3240  
Qy 3241 GATTTCCAACTCTTCCCGCATGTTCTAGTCGCTCTGACTAGACACAAAGAGCTTCAA 3300  
Db 3241 GATTTCCAACTCTTCCCGCATGTTCTAGTCGCTCTGACTAGACACAAAGAGCTTCAA 3300  
Qy 3301 ATATTACACCGTAGTGTAGATCCCTTTAGTACAGATAATTTAGTGTATTTGCTTCTTAAG 3360  
Db 3301 ATATTACACCGTAGTGTAGATCCCTTTAGTACAGATAATTTAGTGTATTTGCTTCTTAAG 3360  
Qy 3361 CTCTCTCTTTTGAATAATGTATGTGTAAGACGAGTAGTAGATAGCAATTTACAGATGGA 3420  
Db 3361 CTCTCTCTTTTGAATAATGTATGTGTAAGACGAGTAGTAGATAGCAATTTACAGATGGA 3420  
Qy 3421 TGCAAGTGTCCAAAGGTCATAATCTCTTTTGTGGCAACACCTTAAATCAGGAGACTTTCCAGA 3480  
Db 3421 TGCAAGTGTCCAAAGGTCATAATCTCTTTTGTGGCAACACCTTAAATCAGGAGACTTTCCAGA 3480  
Qy 3481 TCTACAGTCTTATTACAGATGTATGCTCTCCCTCGGTAAATAGTACTATCTTAAACAAAGTATGA 3540  
Db 3481 TCTACAGTCTTATTACAGATGTATGCTCTCCCTCGGTAAATAGTACTATCTTAAACAAAGTATGA 3540  
Qy 3541 TGCTGTTACATGAGGTTACGTGATAATAGTCTTAAATGTGAGAGGATTTGTTGTTGATTT 3600  
Db 3541 TGCTGTTACATGAGGTTACGTGATAATAGTCTTAAATGTGAGAGGATTTGTTGTTGATTT 3600  
Qy 3601 TTCCAAAGTATTTCCGATGCCAAGGAGGTGAACCATCTAGAGCCAGTTTTCGGTAC 3660  
Db 3601 TTCCAAAGTATTTCCGATGCCAAGGAGGTGAACCATCTAGAGCCAGTTTTCGGTAC 3660  
Qy 3661 CGCGCGGAAACCGCAAGGCTGCAGGACTACTCGAAAAATCTGTTGCAATGATTAAG 3720  
Db 3661 CGCGCGGAAACCGCAAGGCTGCAGGACTACTCGAAAAATCTGTTGCAATGATTAAG 3720  
Qy 3721 AAATTTCAACGCAACGACCTGAGCGGGAAGATTGACATTTGAGAGACCGCATCTGTTGT 3780  
Db 3721 AAATTTCAACGCAACGACCTGAGCGGGAAGATTGACATTTGAGAGACCGCATCTGTTGT 3780  
Qy 3781 AGTAGATAAGTTTTTTCATAGCTATTTTATTTAAAGAGAAAAATACACAAAAATATTGC 3840  
Db 3781 AGTAGATAAGTTTTTTCATAGCTATTTTATTTAAAGAGAAAAATACACAAAAATATTGC 3840  
Qy 3841 TGAGATGATACGCAAGGATTCATATGATGAGATGGTTGAAAAACAGGAAAGATCTATT 3900  
Db 3841 TGAGATGATACGCAAGGATTCATATGATGAGATGGTTGAAAAACAGGAAAGATCTATT 3900  
Qy 3901 GGACGACTTGGCTAACTACAAATTTTACAGATCTCGCGGCCATCGATCGTACAGACAT 3960  
Db 3901 GGACGACTTGGCTAACTACAAATTTTACAGATCTCGCGGCCATCGATCGTACAGACAT 3960  
Qy 3961 GATCAAGGCTCAACCAAAACAGAAATTTGACCTTTCAATTCAGAAATGAATACCTGCTCT 4020  
Db 3961 GATCAAGGCTCAACCAAAACAGAAATTTGACCTTTCAATTCAGAAATGAATACCTGCTCT 4020  
Qy 4021 GCAAACTATTTGCTACCAATTCGAAGCAGATCAACGGTATTTTGGCCGGTTTCTCAGAGCT 4080  
Db 4021 GCAAACTATTTGCTACCAATTCGAAGCAGATCAACGGTATTTTGGCCGGTTTCTCAGAGCT 4080



QY	4081	TACAAGGTGCTGCTCGAGGCAATTTGATCTTAAGAAGTTCTTTCTTTACTAGGAAAC	4140
Db	4081	TACAAGGTGCTGCTCGAGGCAATTTGATCTTAAGAAGTTCTTTCTTTACTAGGAAAC	4140
QY	4141	TCCAGAACAGATTCAGAAATTTTCTCGGATCTCGACTCGCAAGTTCCTATGAGTGT	4200
Db	4141	TCCAGAACAGATTCAGAAATTTTCTCGGATCTCGACTCGCAAGTTCCTATGAGTGT	4200
QY	4201	AGAACTGGATATTTCTAAGTATGATAAGTCCACAGAACGAGTTTCATGTGCTGTAGAT	4260
Db	4201	AGAACTGGATATTTCTAAGTATGATAAGTCCACAGAACGAGTTTCATGTGCTGTAGAT	4260
QY	4261	TGAAATAGGAAAGATTTGGTCTCAATGAGTTTGGCCGAAAGTGTGGAACAAGGGCA	4320
Db	4261	TGAAATAGGAAAGATTTGGTCTCAATGAGTTTGGCCGAAAGTGTGGAACAAGGGCA	4320
QY	4321	CAGGAAACAACCTTTGGAAGATTACATGCTGGAATCAAGACATGCTGTGATCAAA	4380
Db	4321	CAGGAAACAACCTTTGGAAGATTACATGCTGGAATCAAGACATGCTGTGATCAAA	4380
QY	4381	GAAAAGCGGTGATGTGACTACTTTTCATCGGCAATCTGTTATATAGACAGTTCCTTGG	4440
Db	4381	GAAAAGCGGTGATGTGACTACTTTTCATCGGCAATCTGTTATATAGACAGTTCCTTGG	4440
QY	4441	TTCAATGTTACCGATCGAAAGTCAATAAAGTGCTTTTGTGGAGACGATTCGGTTT	4500
Db	4441	TTCAATGTTACCGATCGAAAGTCAATAAAGTGCTTTTGTGGAGACGATTCGGTTT	4500
QY	4501	GTAATTTTCCAAAGGTTTGGATTTCCCTGACATTCAGTCAATGCTTAATCTCATGTGAA	4560
Db	4501	GTAATTTTCCAAAGGTTTGGATTTCCCTGACATTCAGTCAATGCTTAATCTCATGTGAA	4560
QY	4561	TTTTGAGGCCAACTGTATAGAAAGAGTACGTTTACTTTTGTGGTAGATACATACACA	4620
Db	4561	TTTTGAGGCCAACTGTATAGAAAGAGTACGTTTACTTTTGTGGTAGATACATACACA	4620
QY	4621	CCATGATAGAGGCAATAGTGTATATGATCCTTTGAAGTTGATCTCCTCAACTTGGGG	4680
Db	4621	CCATGATAGAGGCAATAGTGTATATGATCCTTTGAAGTTGATCTCCTCAACTTGGGG	4680
QY	4681	AAAAACATATCAAGGATTTATGATCACTTAGAAGATTAAAGGTGCTTTTGTGCGATGTC	4740
Db	4681	AAAAACATATCAAGGATTTATGATCACTTAGAAGATTAAAGGTGCTTTTGTGCGATGTC	4740
QY	4741	TTGTTGCTCGAACTGGTCTTAGGCTTTCCGAGCTGAAACGAGCTATCAAGGAGGT	4800
Db	4741	TTGTTGCTCGAACTGGTCTTAGGCTTTCCGAGCTGAAACGAGCTATCAAGGAGGT	4800
QY	4801	TCATAAAACCGCGATTGATGTTGCTTTGCTTTTAAATGCTTAACAAATTTTGTGTGA	4860
Db	4801	TCATAAAACCGCGATTGATGTTGCTTTGCTTTTAAATGCTTAACAAATTTTGTGTGA	4860
QY	4861	TAAATTTTATTTAGAACTTTGTTTAAATGGCTTTAGTCTCAGAGATCTGTCAAAA	4920
Db	4861	TAAATTTTATTTAGAACTTTGTTTAAATGGCTTTAGTCTCAGAGATCTGTCAAAA	4920
QY	4921	TTAGCGAGTTCATTTGATCTTTCGAAACAGGATGAGATCTTCCGGCATTCATGCTAAG	4980
Db	4921	TTAGCGAGTTCATTTGATCTTTCGAAACAGGATGAGATCTTCCGGCATTCATGCTAAG	4980
QY	4981	TCAAGAGTGTAGAAATATCGACTGTGGAACAAGATTATGGCTGTTAAGAAATGATAGTCT	5040
Db	4981	TCAAGAGTGTAGAAATATCGACTGTGGAACAAGATTATGGCTGTTAAGAAATGATAGTCT	5040
QY	5041	CTGATGTAGATTTACTTAAAGGTTTAAAGTTAGTTAAGAAAGGATGTGTGCTTAGCTG	5100
Db	5041	CTGATGTAGATTTACTTAAAGGTTTAAAGTTAGTTAAGAAAGGATGTGTGCTTAGCTG	5100
QY	5101	ATTTGGTAGTCTTGGGAGTGGAAATCTCCCGGATTAATCTCCGCTGGTGTGCTAGTCT	5160
Db	5101	ATTTGGTAGTCTTGGGAGTGGAAATCTCCCGGATTAATCTCCGCTGGTGTGCTAGTCT	5160
QY	5161	GTAATGTAGATAAGAGAAATGAAGAGTAAAGAGCAACGCTGGGTGCTATCAACGCC	5220

Db	5161	GTAATGTAGATAAGAGAAATGAAGAGTAAAGAGCAACGCTGGGTGCTATCAACGCC	5220
QY	5221	CTGCTTGCAGAAAGAAATTTTCTTTTAAGCTAAATCCCTAAATTTATTCATAATCAATCCGAGG	5280
Db	5221	CTGCTTGCAGAAAGAAATTTTCTTTTAAGCTAAATCCCTAAATTTATTCATAATCAATCCGAGG	5280
QY	5281	ATGCTGAGAGCAACCCGTTGGCAAGTGTAGTGAATATCAAAAGGAGTGGCTATGGAAGAAG	5340
Db	5281	ATGCTGAGAGCAACCCGTTGGCAAGTGTAGTGAATATCAAAAGGAGTGGCTATGGAAGAAG	5340
QY	5341	GATACGTCTCTTTATCTTTTGGAGTTCGTTTCAATTTGTGTAGTACATAAATAATATGTAA	5400
Db	5341	GATACGTCTCTTTATCTTTTGGAGTTCGTTTCAATTTGTGTAGTACATAAATAATATGTAA	5400
QY	5401	GAAAAGTTTTCAGGGAACGTAATTTTGTAGTGTGACAGACGGCTCCGCAATTGAACTCACTG	5460
Db	5401	GAAAAGTTTTCAGGGAACGTAATTTTGTAGTGTGACAGACGGCTCCGCAATTGAACTCACTG	5460
QY	5461	AAAAGGTTTTCAGGAGTTCGTTGATGAAGTACCAATGGCTGTGAAAATCTGAAAAGGTTT	5520
Db	5461	AAAAGGTTTTCAGGAGTTCGTTGATGAAGTACCAATGGCTGTGAAAATCTGAAAAGGTTT	5520
QY	5521	CGGAAAACAAAAGAAATGCTAGCTAATTAATGTTTAATAATAAGAAATAAATAACAGTG	5580
Db	5521	CGGAAAACAAAAGAAATGCTAGCTAATTAATGTTTAATAATAAGAAATAAATAACAGTG	5580
QY	5581	GTAAGAGGGTTTAAATTTGAGGAAATGAGGAAATTAAGTGTGATGACGAGTCTATCG	5640
Db	5581	GTAAGAGGGTTTAAATTTGAGGAAATGAGGAAATTAAGTGTGATGACGAGTCTATCG	5640
QY	5641	CGTCATCGAGTACGTTTAAATCAATGATCCCTTATACAACTCACTCCGAGCCAAATTTGT	5700
Db	5641	CGTCATCGAGTACGTTTAAATCAATGATCCCTTATACAACTCACTCCGAGCCAAATTTGT	5700
QY	5701	TTACTTATCTTCCGCTTACGACAGTCTGTGACAGCTGATCAATCTGTGTACAAATGCATT	5760
Db	5701	TTACTTATCTTCCGCTTACGACAGTCTGTGACAGCTGATCAATCTGTGTACAAATGCATT	5760
QY	5761	GGGTAAACAGTTTCAACAGCAACAGCTAGGACAAACAGTCCAAACAGCAATTTCCGGATGC	5820
Db	5761	GGGTAAACAGTTTCAACAGCAACAGCTAGGACAAACAGTCCAAACAGCAATTTCCGGATGC	5820
QY	5821	CTGAAAACCTGTGCTAGTATGACAGTGAATTTCTGCACTCGATTTCTATGTGTATAG	5880
Db	5821	CTGAAAACCTGTGCTAGTATGACAGTGAATTTCTGCACTCGATTTCTATGTGTATAG	5880
QY	5881	ATATAATTCAGCTTGTATCCGTTGATCAACGCGGTTTAAATAGCTTTGATACATAA	5940
Db	5881	ATATAATTCAGCTTGTATCCGTTGATCAACGCGGTTTAAATAGCTTTGATACATAA	5940
QY	5941	TAGAATAATAGAGTTTGAATAATCAACCCGACCGAATCTACTGAAATCGTTAAACCGGAC	6000
Db	5941	TAGAATAATAGAGTTTGAATAATCAACCCGACCGAATCTACTGAAATCGTTAAACCGGAC	6000
QY	6001	TCAGAGGTAGACGATGCTACTTATAGGCTTCAATCAATAATTTTGGCTAATGA	6060
Db	6001	TCAGAGGTAGACGATGCTACTTATAGGCTTCAATCAATAATTTTGGCTAATGA	6060
QY	6061	ACTGGTTCCGTGGAACTGGCATGTTCAATCAAGCAGCTTTGAGACTGTGTGGACTTGT	6120
Db	6061	ACTGGTTCCGTGGAACTGGCATGTTCAATCAAGCAGCTTTGAGACTGTGTGGACTTGT	6120
QY	6121	CTGGACCAACAATCTCCGGCTACTTAGCTATTTGTGTGAGATTTCTTAAATAAAGTCGCTG	6180
Db	6121	CTGGACCAACAATCTCCGGCTACTTAGCTATTTGTGTGAGATTTCTTAAATAAAGTCGCTG	6180
QY	6181	AGAATTTAAATTCAGGGTGGCTGATACCAAAATCAGCAGTGGTGTTCCTCCACTTAAA	6240
Db	6181	AGAATTTAAATTCAGGGTGGCTGATACCAAAATCAGCAGTGGTGTTCCTCCACTTAAA	6240
QY	6241	TATAACGATTTGTCATATCTGGATCCAAACAGTTTAAACCAATGTGATGCTGTATCTGGTA	6300



Db 6241 TATAACGATTGTCTATATCTCGATCAACAGATTAAACCAATGTGATGGTGTATCTGTGTA 6300  
Qy 6301 TGGCGTAAACATCGGAGAGTTGGAATCTCCCTAACCGCGGTAGCGGCCA 6355  
Db 6301 TGGCGTAAACATCGGAGAGTTGGAATCTCCCTAACCGCGGTAGCGGCCA 6355  
RESULT 2  
ID ACC85005 standard; DNA; 6355 BP.  
AC ACC85005;  
XX 13-OCT-2003 (first entry)  
DT TMV-U2 genome nucleotide sequence.  
DE TMV-U2 genome nucleotide sequence.  
XX Inhibitory RNA; viral RNA vector; coat protein; TMV; U2; gene; ds.  
XX Tobacco mosaic virus.  
XX WO2003052108-A2.  
XX 26-JUN-2003.  
XX 05-DEC-2002; 2002WO-EP013964.  
XX 18-DEC-2001; 2001US-0340488P.  
XX (FARB ) BAYER BIOSCIENCE NV.  
XX Metzlaiff MH, Gossele WML, Meulewaeter F, Fache ICA;  
XX WPI; 2003-523529/49.  
XX Introducing inhibitory RNA into a plant cell comprises providing a viral  
XX RNA vector derived from a satellite RNA virus that encodes a coat  
XX protein, and infecting a plant with the viral RNA vector and a  
XX corresponding helper virus.  
XX Example; Page 79-82; 86pp; English.  
XX The invention relates to introducing inhibitory RNA into a plant cell.  
XX The method involves providing a viral RNA vector derived from a satellite  
XX RNA virus having a sequence that encodes a coat protein, and infecting a  
XX plant with the viral RNA vector and a corresponding helper virus. The  
XX methods and viral RNA vectors are useful in introducing inhibitory RNA  
XX into plant cells. These may be used to determine or validate the function  
XX of isolated nucleic acid sequences in plants. The present sequence  
XX represents the nucleotide sequence of the genome of tobacco mosaic virus  
XX (TMV) -U2  
SQ Sequence 6355 BP; 1933 A; 1112 C; 1489 G; 1821 T; 0 U; 0 Other;  
Query Match 100.08; Score 6355; DB 9; Length 6355;  
Best Local Similarity 100.08; Pred. No. 0;  
Matches 6355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GATGTTTTTAATAGTTTTCGACAAACAAATTAATAACAAACAAACATATTACAAACACA 60  
Db 1 GATGTTTTTAATAGTTTTCGACAAACAAATTAATAACAAACAAACATATTACAAACACA 60  
Qy 61 AACAAACAATGGCACACATACATCTATTAATAGCAACGCCCTTCTTGAAGCGTGAG 120  
Db 61 AACAAACAATGGCACACATACATCTATTAATAGCAACGCCCTTCTTGAAGCGTGAG 120  
Qy 121 TGGTAAACAACTCTCGTTAATGACCTTGCAGAGGCGCATGTACGATCGCGCGTGA 180  
Db 121 TGGTAAACAACTCTCGTTAATGACCTTGCAGAGGCGCATGTACGATCGCGCGTGA 180  
Qy 181 AGAATTTAACCGCGCGCGGTAGCAACAAAGTCAACTTTTCCAAACCTATTAGCGAAGA 240  
Db 181 AGAATTTAACCGCGCGCGGTAGCAACAAAGTCAACTTTTCCAAACCTATTAGCGAAGA 240

Qy 241 GCAAACCGCTTCTAGTCTCCAAACGGGTACCCGAGTTCAGANTTACCTTTTATAAFACTCA 300  
Db 241 GCAAACCGCTTCTAGTCTCCAAACGGGTACCCGAGTTCAGANTTACCTTTTATAAFACTCA 300  
Qy 301 AAATGCCGTACACAGTTTGGCTGGAGGTTTGGAGGACATTAGAAATTGGAATATCTGATGCT 360  
Db 301 AAATGCCGTACACAGTTTGGCTGGAGGTTTGGAGGACATTAGAAATTGGAATATCTGATGCT 360  
Qy 361 ACAAGTTCCCTATGCGGACATATGATATATAGTGGGAACTTTGCGAGCACATTTGTT 420  
Db 361 ACAAGTTCCCTATGCGGACATATGATATATAGTGGGAACTTTGCGAGCACATTTGTT 420  
Qy 421 CAAAGGCAAGGATTTACGTGCAATTCCTGTAAGCCCAATCTGACACATACGAGATATAATGAG 480  
Db 421 CAAAGGCAAGGATTTACGTGCAATTCCTGTAAGCCCAATCTGACACATACGAGATATAATGAG 480  
Qy 481 GCACGAAGGACAAAGGACTCAATTTGAGATGTAATTTGTCAGATTTGTCCTGTTCTTAACA 540  
Db 481 GCACGAAGGACAAAGGACTCAATTTGAGATGTAATTTGTCAGATTTGTCCTGTTCTTAACA 540  
Qy 541 GGTAAATTCCTGAGTTTCAAAGGGAGGCTTTTAAACAGGTATGCAGAACTCCCAACGAAGT 600  
Db 541 GGTAAATTCCTGAGTTTCAAAGGGAGGCTTTTAAACAGGTATGCAGAACTCCCAACGAAGT 600  
Qy 601 CTGCTGCTCTAAACCTTTTCAAGGATTTGTCGAATACATCCCGCAGAGATAGTGTAGAG 660  
Db 601 CTGCTGCTCTAAACCTTTTCAAGGATTTGTCGAATACATCCCGCAGAGATAGTGTAGAG 660  
Qy 661 ATACGCTGTTGCTCTGCACAGTTTGTATGATATTTCTGTGTCATGATGTTGAGGCTGCTT 720  
Db 661 ATACGCTGTTGCTCTGCACAGTTTGTATGATATTTCTGTGTCATGATGTTGAGGCTGCTT 720  
Qy 721 AATATCTAAGAATATACATGTATGTTATGCAAGCTTCCATTTTGGCAGAGCATTTATTACT 780  
Db 721 AATATCTAAGAATATACATGTATGTTATGCAAGCTTCCATTTTGGCAGAGCATTTATTACT 780  
Qy 781 AGACGACGAGGATTTAGCTTAATGAATAGGCGCAACTTTCAAAGAGAGAGTGTATGA 840  
Db 781 AGACGACGAGGATTTAGCTTAATGAATAGGCGCAACTTTCAAAGAGAGAGTGTATGA 840  
Qy 841 TGTGTTTTTCTTCTGATGAAAGTACTTTTAAATTTATAGTCATATAACAAAAATAT 900  
Db 841 TGTGTTTTTCTTCTGATGAAAGTACTTTTAAATTTATAGTCATATAACAAAAATAT 900  
Qy 901 CTTGCATTATGTAGTTAAATCTTACTTTCTGCTTCTAGTAGAATAGTTTACTTTAAGGA 960  
Db 901 CTTGCATTATGTAGTTAAATCTTACTTTCTGCTTCTAGTAGAATAGTTTACTTTAAGGA 960  
Qy 961 ATTTTGTAGTCACTAGGTTTAAATCTTGGTTTGTAAATTTACCAAGTAGATACCTATAT 1020  
Db 961 ATTTTGTAGTCACTAGGTTTAAATCTTGGTTTGTAAATTTTACCAAGTAGATACCTATAT 1020  
Qy 1021 TCTGTACAAGAGTCTTAGACAAGTAGGTGTAGTAGTGTATGATCAGTTCTATGAGCGATGA 1080  
Db 1021 TCTGTACAAGAGTCTTAGACAAGTAGGTGTAGTAGTGTATGATCAGTTCTATGAGCGATGA 1080  
Qy 1081 AGACGCTTTGCTTTACAAGAAACCTTTGGCCATGTTTCAACACTGAAGAGCAATCTTTAG 1140  
Db 1081 AGACGCTTTGCTTTACAAGAAACCTTTGGCCATGTTTCAACACTGAAGAGCAATCTTTAG 1140  
Qy 1141 AGACGCGCTTTCGGTTTAACTTTTGGTTCCCTTAAGTAGAAGGACATGGTGTAGTACCCGT 1200  
Db 1141 AGACGCGCTTTCGGTTTAACTTTTGGTTCCCTTAAGTAGAAGGACATGGTGTAGTACCCGT 1200  
Qy 1201 GTTTGAGGGTTCTATTACCAGCAAAAGATGCAAGGAGTCAAGTGTATTTAATCGTGA 1260  
Db 1201 GTTTGAGGGTTCTATTACCAGCAAAAGATGCAAGGAGTCAAGTGTATTTAATCGTGA 1260  
Qy 1261 CTTGCTTTTACACAGTGTCTTAATCATATCAGAACATATCAAGCCAAAGCGTTAACTTACCA 1320  
Db 1261 CTTGCTTTTACACAGTGTCTTAATCATATCAGAACATATCAAGCCAAAGCGTTAACTTACCA 1320

QY 1321 GAAAGTATATCTTTCTGAGGAGTCTATAAGATCCCGCGTGATTAATCAATGGTGTACTGC 1380  
DB 1321 GAAAGTATATCTTTCTGAGGAGTCTATAAGATCCCGCGTGATTAATCAATGGTGTACTGC 1380  
QY 1381 TAGGTCGTGAATGGAGTAGATAAAGCAATCTTCAACCTTGTCAATGACCTTCTTCT 1440  
DB 1381 TAGGTCGTGAATGGAGTAGATAAAGCAATCTTCAACCTTGTCAATGACCTTCTTCT 1440  
QY 1441 GCAGACTAAGCTGGCTGCGCTTCAAGACGATATAGTAATGGGAAGTTCGGTGTCTGGA 1500  
DB 1441 GCAGACTAAGCTGGCTGCGCTTCAAGACGATATAGTAATGGGAAGTTCGGTGTCTGGA 1500  
QY 1501 TAAGACCACTTCTGAATCTTATTTGGGATGAGTGGGCAAAATTTTGGAAAAGTTCCTCC 1560  
DB 1501 TAAGACCACTTCTGAATCTTATTTGGGATGAGTGGGCAAAATTTTGGAAAAGTTCCTCC 1560  
QY 1561 CACTATCAAGAGAGATTCGTGAGCAGGAAATCTCGGATGTAAGTGAGAAATGCTCTGAA 1620  
DB 1561 CACTATCAAGAGAGATTCGTGAGCAGGAAATCTCGGATGTAAGTGAGAAATGCTCTGAA 1620  
QY 1621 GATCAAGATCCAGATCTGTATGTCCATGGAAGACAGGTTCTGTAGCTGAATACACCAA 1680  
DB 1621 GATCAAGATCCAGATCTGTATGTCCATGGAAGACAGGTTCTGTAGCTGAATACACCAA 1680  
QY 1681 GTCTGAGGAGTTACCGCATCTAGATATCAAGAAAGCTTAGAAGAGCTGAGCAAAATGTA 1740  
DB 1681 GTCTGAGGAGTTACCGCATCTAGATATCAAGAAAGCTTAGAAGAGCTGAGCAAAATGTA 1740  
QY 1741 CGACGCGTTATCAGAAATCTATCTTAAAGGTCGTGATATTTTGGATTCGCAAGTT 1800  
DB 1741 CGACGCGTTATCAGAAATCTATCTTAAAGGTCGTGATATTTTGGATTCGCAAGTT 1800  
QY 1801 CAAAGACATGTCAGAGCTTTAGATGTTAGTCTCTGATGTGGCAGCACAGTAATCTGTGC 1860  
DB 1801 CAAAGACATGTCAGAGCTTTAGATGTTAGTCTCTGATGTGGCAGCACAGTAATCTGTGC 1860  
QY 1861 AGTGCCGAGATAGACGGTTTAACTCTTACTTTTGTATAGCCAAACCGAGGAGATGT 1920  
DB 1861 AGTGCCGAGATAGACGGTTTAACTCTTACTTTTGTATAGCCAAACCGAGGAGATGT 1920  
QY 1921 GCGTAAAGCTCTTAAAGCACGCGCTCTGAGGCGGTGTATGTTTGAACCGACATCCGA 1980  
DB 1921 GCGTAAAGCTCTTAAAGCACGCGCTCTGAGGCGGTGTATGTTTGAACCGACATCCGA 1980  
QY 1981 AGAGTGAACTGTAATAATTTCTATGCTGAGAAAGGAGATGTCCTGTGTGTGCAGA 2040  
DB 1981 AGAGTGAACTGTAATAATTTCTATGCTGAGAAAGGAGATGTCCTGTGTGTGCAGA 2040  
QY 2041 AAGTCATGTTTGACGAATGCTTAACTAGACACCGAGGTTGGAGTCCCTCAACGATTT 2100  
DB 2041 AAGTCATGTTTGACGAATGCTTAACTAGACACCGAGGTTGGAGTCCCTCAACGATTT 2100  
QY 2101 CCATAGGCTTTCGCTGAGTAGTGTGATTAAGCAAAATGGCATCGGTTGCTACACTGG 2160  
DB 2101 CCATAGGCTTTCGCTGAGTAGTGTGATTAAGCAAAATGGCATCGGTTGCTACACTGG 2160  
QY 2161 CTCACCTAAAGTTTCAAAATGAGAACTATGTGGAAGATTTGGCAGCTTCGTTGTCGC 2220  
DB 2161 CTCACCTAAAGTTTCAAAATGAGAACTATGTGGAAGATTTGGCAGCTTCGTTGTCGC 2220  
QY 2221 CACTGTATCAAACTTATCAAGTCACTAAAGATGAAGTCCGGTATGATTCGTATCCAG 2280  
DB 2221 CACTGTATCAAACTTATCAAGTCACTAAAGATGAAGTCCGGTATGATTCGTATCCAG 2280  
QY 2281 GGAGAAAGTTGGTGTGGGATGTCATTTGAAAAGTGGCTCCCTCAAACTCGGGGCAA 2340  
DB 2281 GGAGAAAGTTGGTGTGGGATGTCATTTGAAAAGTGGCTCCCTCAAACTCGGGGCAA 2340  
QY 2341 AGGTCATTCATGGGAGTTGTCCTGGATTAAGAGGGAATGTTTACTGACCTTCTATC 2400  
DB 2341 AGGTCATTCATGGGAGTTGTCCTGGATTAAGAGGGAATGTTTACTGACCTTCTATC 2400  
QY 2401 TTATGAAGGAGATAGAATGGTGACTGAGACGCACTGGAGGAGGCTGTATCATCTGA 2460

DB 2401 TTATGAAGGAGATAGAATGGTGACTGAGACGCACTGGAGGAGGCTGTATCATCTGA 2460  
QY 2461 TACAATGATATATCTTGATATTTGAAAGCTCCAAATCTGAGGAAAAAACAATGAGACGG 2520  
DB 2461 TACAATGATATATCTGATATTTGAAAGCTCCAAATCTGAGGAAAAAACAATGAGACGG 2520  
QY 2521 TGAACCCACCAACCTACTGCAAGATGATTTGATCTTGAAGATTTGATCTTGGTTCCTGGAAA 2580  
DB 2521 TGAACCCACCAACCTACTGCAAGATGATTTGATCTTGAAGATTTGATCTTGGTTCCTGGAAA 2580  
QY 2581 GTACAAAGGAGATTTTGAAGATTTGATCTTGAAGATTTGATCTTGGTTCCTGGAAA 2640  
DB 2581 GTACAAAGGAGATTTTGAAGATTTGATCTTGAAGATTTGATCTTGGTTCCTGGAAA 2640  
QY 2641 ACAAGCTGCTGTATGATCAGAAAGGCTTAATTCATCTGGAATGATGAGGACCAAT 2700  
DB 2641 ACAAGCTGCTGTATGATCAGAAAGGCTTAATTCATCTGGAATGATGAGGACCAAT 2700  
QY 2701 GGACAAATGTGAGAACGGTAGATTCACTTCTAATCATCCAAACCGGATCACACAGAG 2760  
DB 2701 GGACAAATGTGAGAACGGTAGATTCACTTCTAATCATCCAAACCGGATCACACAGAG 2760  
QY 2761 GCTTTTATGATGAAGGTTGATGCTGCACACACCGTGTGTAACTTCTCTGGTCTTAT 2820  
DB 2761 GCTTTTATGATGAAGGTTGATGCTGCACACACCGTGTGTAACTTCTCTGGTCTTAT 2820  
QY 2821 CTCTGGTTGCGACATCCATACATTTAGGAGATACACAGCAGATTCCTTTCATTAACAG 2880  
DB 2821 CTCTGGTTGCGACATCCATACATTTAGGAGATACACAGCAGATTCCTTTCATTAACAG 2880  
QY 2881 AGTTTCAAGATTTCCGCTATCCAAACATTTTGAAGCTGCAAGTGGATGAAGTTGAGAT 2940  
DB 2881 AGTTTCAAGATTTCCGCTATCCAAACATTTTGAAGCTGCAAGTGGATGAAGTTGAGAT 2940  
QY 2941 GAGGAGACCACTGATGATGCTCCAGGTGATGTAATTTTCTCAATGAAAGTACGA 3000  
DB 2941 GAGGAGACCACTGATGATGCTCCAGGTGATGTAATTTTCTCAATGAAAGTACGA 3000  
QY 3001 AGGAGCGGTGACACCACTTCACTGTACACAGATCGGTCTCATCTGAGATGATAGCGG 3060  
DB 3001 AGGAGCGGTGACACCACTTCACTGTACACAGATCGGTCTCATCTGAGATGATAGCGG 3060  
QY 3061 TAAGGAGTAGTAAACAGTGTTCCTCAACCACTAAAAAGGAAAAATTTGACTTCACTCA 3120  
DB 3061 TAAGGAGTAGTAAACAGTGTTCCTCAACCACTAAAAAGGAAAAATTTGACTTCACTCA 3120  
QY 3121 GGCTGATAAATTTGAGTTAGGAGGAGGCTATAAGAAATGTGAAACACCGTTCTATGAGAT 3180  
DB 3121 GGCTGATAAATTTGAGTTAGGAGGAGGCTATAAGAAATGTGAAACACCGTTCTATGAGAT 3180  
QY 3181 CCAAGGAGAAAACTTTGAGATGTCGCTGGTCAGATTCACGCACTCCACTGACTCT 3240  
DB 3181 CCAAGGAGAAAACTTTGAGATGTCGCTGGTCAGATTCACGCACTCCACTGACTCT 3240  
QY 3241 GATTTCCAGTCTTCCCGCATGTTCTAGTCTGCTGACTAGACACAAAGAGCTTCAA 3300  
DB 3241 GATTTCCAGTCTTCCCGCATGTTCTAGTCTGCTGACTAGACACAAAGAGCTTCAA 3300  
QY 3301 ATATTACACCGTAGTGTAGATCCTTTAGTACAGATTAATTTAGTGTCTTCTTTAAG 3360  
DB 3301 ATATTACACCGTAGTGTAGATCCTTTAGTACAGATTAATTTAGTGTCTTCTTTAAG 3360  
QY 3361 CTCTCTCTTTTGAAGATGATATGTTAGAACGAGGTAGTAGATAGCAATTAAGATGGA 3420  
DB 3361 CTCTCTCTTTTGAAGATGATATGTTAGAACGAGGTAGTAGATAGCAATTAAGATGGA 3420  
QY 3421 TGCAGTGTTCAAAGGTCATTAATCTCTTTGTGGCAACCTTAAATCAGGAGACTTCCAGA 3480  
DB 3421 TGCAGTGTTCAAAGGTCATTAATCTCTTTGTGGCAACCTTAAATCAGGAGACTTCCAGA 3480  
QY 3481 TCTACAGTTCATTAAGTATGCTGCTCCCTGCTTAATAGTACTATTAACAGATAGA 3540

3481	TCTACAGTTCCTATTACGATGTATGCCCTCCCTGGTAATAGTACTATATCTTAAACAAGTATGA	3541	TGCTGTTTACCATTGAGGTTTACGTGATATAGTCTTAAATGTGAAGGATGTGTTCTTCGATTT
3541	TGCTGTTTACCATTGAGGTTTACGTGATATAGTCTTAAATGTGAAGGATGTGTTCTTCGATTT	3601	TTCCAAAAGATATTTCCGATGCGAAAGGAGGTGAAACCAATGCTTAGAGCCAGTTTGTGGGTAC
3601	TTCCAAAAGATATTTCCGATGCGAAAGGAGGTGAAACCAATGCTTAGAGCCAGTTTGTGGGTAC	3661	CGCGGCGGAACGCCCAAGGGCTCGAGACTACTCGRAAAATCTGGTTGCAATGATATAAAG
3661	CGCGGCGGAACGCCCAAGGGCTCGAGACTACTCGRAAAATCTGGTTGCAATGATATAAAG	3721	AAATTTCAACCGCACCCAGACCTCGACGGGACGATTTGACATTTGAGACGACCCGCATCTGTTGT
3721	AAATTTCAACCGCACCCAGACCTCGACGGGACGATTTGACATTTGAGACGACCCGCATCTGTTGT	3781	AGTAGATAAGTTTTTTTGATAGCTATTTTTATTTAAAGAAAGAAATATACAAAAAATATTCG
3781	AGTAGATAAGTTTTTTTGATAGCTATTTTTATTTAAAGAAAGAAATATACAAAAAATATTCG	3841	TGGAGTGATGACGAAGGATTTCAATGATGAGATGGTTGGRAAAACAGNAAGAAGTACTATT
3841	TGGAGTGATGACGAAGGATTTCAATGATGAGATGGTTGGRAAAACAGNAAGAAGTACTATT	3901	GGACGACTTGGCTAACTTACAAATTTTACAGATCTGCCGGCCCATCGATCAGTACAAAGCACAT
3901	GGACGACTTGGCTAACTTACAAATTTTACAGATCTGCCGGCCCATCGATCAGTACAAAGCACAT	3961	GATCAAGGCTCAACCAAAAACAGAAATTTGGACCTTTTCAATTCAGAAATACCCCTGCTCT
3961	GATCAAGGCTCAACCAAAAACAGAAATTTGGACCTTTTCAATTCAGAAATACCCCTGCTCT	4021	GCAAAACAATTTGTCTACCATTTGCAAGCAGATCAACGGTATTTTGGCCGGTTTCTCAGAGCT
4021	GCAAAACAATTTGTCTACCATTTGCAAGCAGATCAACGGTATTTTGGCCGGTTTCTCAGAGCT	4081	TACAAGGTTGCTGCTCGAGGCATTTGATTTCTAAGAGTTTCTTTTCTTTACTAGGAAAC
4081	TACAAGGTTGCTGCTCGAGGCATTTGATTTCTAAGAGTTTCTTTTCTTTACTAGGAAAC	4141	TCCAGAACAGATTTCAAGAAATTTTTTCTCGGATCTCGACTCGCACGTTCTTATGGAATGTGT
4141	TCCAGAACAGATTTCAAGAAATTTTTTCTCGGATCTCGACTCGCACGTTCTTATGGAATGTGT	4201	AGAACTCGGATATTTCTAAGTATGATAGTCAAGAACGAGTTTCATTTGCTGCTGATAGATG
4201	AGAACTCGGATATTTCTAAGTATGATAGTCAAGAACGAGTTTCATTTGCTGCTGATAGATG	4261	TGAAAATATGAAAAGATTTGGGTCTCAATGAGTTTTTGGCCGAAGTGTGGAAAACAAGGCA
4261	TGAAAATATGAAAAGATTTGGGTCTCAATGAGTTTTTGGCCGAAGTGTGGAAAACAAGGCA	4321	CAGGAAAAACAATTTTGAAGGATTTACATTTGCTGGAATCAAGACATGTCGTGGGTATCAAAG
4321	CAGGAAAAACAATTTTGAAGGATTTACATTTGCTGGAATCAAGACATGTCGTGGGTATCAAAG	4381	GAAAAGCGGTGATGTGACTACTTTTCAATCGGCAATTAATAGCAGCTGCTGTTGGG
4381	GAAAAGCGGTGATGTGACTACTTTTCAATCGGCAATTAATAGCAGCTGCTGTTGGG	4441	TTCAATGTTTACCGATGGAAGGTCATAAAGGTGCTTTTTTGTGGAGACCATTCGGTTTT
4441	TTCAATGTTTACCGATGGAAGGTCATAAAGGTGCTTTTTTGTGGAGACCATTCGGTTTT	4501	GTATTTTTCCAAAAGGTTTGGATTTCCCTGCATTTCAATGCTGCTAATCTCATGTGGAA
4501	GTATTTTTCCAAAAGGTTTGGATTTCCCTGCATTTCAATGCTGCTAATCTCATGTGGAA	4561	TTTTGAGGCCAAACTGTATAGAAAGAGGTACGGTTACTTTTTGTGGTATAGATACATCATACA
4561	TTTTGAGGCCAAACTGTATAGAAAGAGGTACGGTTACTTTTTGTGGTATAGATACATCATACA		

Qy		4621	CCATGATAAGGGAGCAATAGTGTATTATGATCTCTTTGAAGTTTGATCTCMAAACCCTCGGGC	4680
Db		4621	CCATGATAAGGGAGCAATAGTGTATTATGATCTCTTTGAAGTTTGATCTCMAAACCCTCGGGC	4680
Qy		4681	AAAACATATCAAGATTTATGATCACTTAGAAGAGTTTAAGGGTGTCTTTGTGCGATGTTC	4740
Db		4681	AAAACATATCAAGATTTATGATCACTTAGAAGAGTTTAAGGGTGTCTTTGTGCGATGTTC	4740
Qy		4741	TGTGTCCTCGGAACCTGGTGCTTAGGCTTCCGACAGCTGAACGCAGCTATCAAGAGGT	4800
Db		4741	TGTGTCCTCGGAACCTGGTGCTTAGGCTTCCGACAGCTGAACGCAGCTATCAAGAGGT	4800
Qy		4801	TCATAAAACCGCATGTGATGGTTTCGTTTCTTTTAAATGTTTAAACAATTTTTTGTGTGA	4860
Db		4801	TCATAAAACCGCATGTGATGGTTTCGTTTCTTTTAAATGTTTAAACAATTTTTTGTGTGA	4860
Qy		4861	TAAATTTTTTATTAGAACCTTTGTTTTTAAATGGCTGTAGTCTCAGAGATACTGTCAAAA	4920
Db		4861	TAAATTTTTTATTAGAACCTTTGTTTTTAAATGGCTGTAGTCTCAGAGATACTGTCAAAA	4920
Qy		4921	TTAGCGAGTCAATTGATCTTTTCAAACAGGATGAGATACCTCCGGCATTCATGACTAAGG	4980
Db		4921	TTAGCGAGTCAATTGATCTTTTCAAACAGGATGAGATACCTCCGGCATTCATGACTAAGG	4980
Qy		4981	TCAAGAGTGTAGAATAATCGACTGTGCACAAGATTAATGGCTGTTTAAGAAATGATAGTCTTT	5040
Db		4981	TCAAGAGTGTAGAATAATCGACTGTGCACAAGATTAATGGCTGTTTAAGAAATGATAGTCTTT	5040
Qy		5041	CTGATGTAGATTACTTAAAGGTGTTAAGTTAGTTAAGAAAGGTTATGTGCTTAGCTG	5100
Db		5041	CTGATGTAGATTACTTAAAGGTGTTAAGTTAGTTAAGAAAGGTTATGTGCTTAGCTG	5100
Qy		5101	ATTGTTAGTGTCTGGGAGTGAATCTCCGATTAACCTGCCGTGTGGTGTCAAGTGTTT	5160
Db		5101	ATTGTTAGTGTCTGGGAGTGAATCTCCGATTAACCTGCCGTGTGGTGTCAAGTGTTT	5160
Qy		5161	GTAATTGTAGATAAGAGAAATCAAAGGAGTAAGGAAGCAACGCTGGGTGCGTATCACGCC	5220
Db		5161	GTAATTGTAGATAAGAGAAATCAAAGGAGTAAGGAAGCAACGCTGGGTGCGTATCACGCC	5220
Qy		5221	CTGCTTCCAAAAAGAAATTTTTCTTTTAAGCTAATCCCTAATTAATCAATTAACATCCGAGG	5280
Db		5221	CTGCTTCCAAAAAGAAATTTTTCTTTTAAGCTAATCCCTAATTAATCAATTAACATCCGAGG	5280
Qy		5281	ATGCTGAGAAGCACCGGTGCAAGTGTTAGTGAATATCAAAGGAGTGGCTATCGAAGAAG	5340
Db		5281	ATGCTGAGAAGCACCGGTGCAAGTGTTAGTGAATATCAAAGGAGTGGCTATCGAAGAAG	5340
Qy		5341	GATACCTCTCTTATCTTTGGAGTTCGTTTCAATTTGTGTAGTACATAAAAAATAATGTAA	5400
Db		5341	GATACCTCTCTTATCTTTGGAGTTCGTTTCAATTTGTGTAGTACATAAAAAATAATGTAA	5400
Qy		5401	GAANAAGTTTGGGGAACGTAATTTTCAGTGTGACAGCGCTCCGCCAATTTGAACCTCACTG	5460
Db		5401	GAANAAGTTTGGGGAACGTAATTTTCAGTGTGACAGCGCTCCGCCAATTTGAACCTCACTG	5460
Qy		5461	AAAAGTGTGTGAGGAGTTCGTGGATGAAGTACCAATGGCTGTGAAACTCGAANAAGTTTC	5520
Db		5461	AAAAGTGTGTGAGGAGTTCGTGGATGAAGTACCAATGGCTGTGAAACTCGAANAAGTTTC	5520
Qy		5521	CGHAACAAAAAGAAATGGTAGGTAAATGTTTAATTAATAGAAATAATAATCAAGTG	5580
Db		5521	CGHAACAAAAAGAAATGGTAGGTAAATGTTTAATTAATAGAAATAATAATCAAGTG	5580
Qy		5581	GTAAGAAGGTTTTTAAANNTTGAAGAAATGAGGATTAATGTAAGTGATGACGAGTCTATCG	5640
Db		5581	GTAAGAAGGTTTTTAAANNTTGAAGAAATGAGGATTAATGTAAGTGATGACGAGTCTATCG	5640
Qy		5641	CGTCAATCGAGTACGTTTTTAATCAATATGCTTATACAATCAACTCTCCGAGCCAAATTTGT	5700
Db		5641	CGTCAATCGAGTACGTTTTTAATCAATATGCTTATACAATCAACTCTCCGAGCCAAATTTGT	5700



Qy	558	AAAGGGAGGCTTTTAAACAGGTATGCAGAGCTCCCAACGAAGTCTGCTGTCTTAAACATT	617
Db	559	AAAAGGAAGCTTTTCGACAGATACCGCTGAAATGCCAAACGAAGTAGTCTGTCAAGTACTT	618
Qy	618	TTCAGGATTCTCGAATACATCCGCCAGAGAAATAGTGTAGAGATACGCTGTGTCTCTGC	677
Db	619	TCCAAAGTGTAGGCAATCTCAAGATGTTCACGGNAGAGTGTATGCTATTGCTTTGCG	678
Qy	678	ACAGTTTGTATGATPATTTCTGTGTCAGTAGTTTGGAGCTGCGTTAAATATCTAAGAAATATAC	737
Db	679	ATAGTATATACGATATACCTGCCGACAGTTCCGCGCGCACTGCTGAGAAAGAAATGTAC	738
Qy	738	ATGATATGTTATGCAAGCTTCCATTTTGGCAGAGCAATTAATTACTAGACGACGAGGTTA	797
Db	739	ATGTATGTTATGCGCGCTTCCACTTTTCCGAGAATTTACTTCTCGAAGATTCACACGTCA	798
Qy	798	CGCTTAATGAAATAGGCGCAACTTTCAAAGAGAGAGTGTATGTTCTTTTCTTTCTTTG	857
Db	799	ACCTCGACGAGATCATATGCTATGTTTCCAAAGAGATGGAGACAGTTGACATTTTCCTTTG	858
Qy	858	CTGATGAAAGTACTTTTAAATTTATAGTCATAAATACAAAAAATATCTTGCAATTGTAGTTA	917
Db	859	CATCTGAGAGTACTTTAAATATAGTCATAGTTATTCTTAATATCTTAAATATGTTTGCA	918
Qy	918	AACTCTACTTTCGTGCTCTAGTAGAATAATGTTACTTTTAAAGAAATTTTTTAGTCACATAGG	977
Db	919	AAACTTACTTCCCAAGCCTCTAAATAGAGAGGTTTACATGAAGGAGTTTTTATAGTAACATAG	978
Qy	978	TTAATACTTGGTTTTGTAAATTTTACCAAGTAGNACTATATCTCTGTACAAGAGTGTTA	1037
Db	979	TTAATACTCGTTTTGTAAATTTTCTAGAATAGATACTTTCTTATTGTACAAGGTGTAG	1038
Qy	1038	GACAAGTAGGCTGTGATAGTAGTCAAGTTCATATGAGCGCATGGAAGACGCTTTCGCTTACA	1097
Db	1039	CGCATAGGGTGTAGATAGTGAGCAGTTTTTACAAGGCTATGGAAGACGATGGCACTACA	1098
Qy	1098	AGAAACCTTTGGCCATGTTCAACACTGAAAGAGCAATCTTTTAGAGACACGCGCTTCGGTTA	1157
Db	1099	AAAAGACTCTTCGATGTGCAACAGTGAAGAAATCTTTGTTAGAGGATCTTTCATCAGTTA	1158
Qy	1158	ACTTTTGGTTCCCTTAAGATGAAGACATGGTGATAGTACCGCTGTTTGGGGTTCATTA	1217
Db	1159	ATTACTCGTTTCCAAAAATAGGGGATATGGTGATGTTCCACTATTTGACATATCTCTCG	1218
Qy	1218	CCAGCAAAAGATGCAACAGGAGTGAGGTCAATGTGTAATCGTGACTTCGTTTACACAGTGC	1277
Db	1219	AGACTAGTAAAGAACACGCNAGAGGCTTAGTTTCAAGGACTTTGTTTATACAGTGT	1278
Qy	1278	TTAATCATATCAGAACATATCAAGCCAAAGCGTTAACTTTACGAAACGTTATATCTTTTCG	1337
Db	1279	TAAATCATACTTCGTACGTACAGGCCAAAGCGCTTACTTACTCCAAAGTGTATCTTTTCG	1338
Qy	1338	TGGAGTCTATAAGATCCCGGTGTATTAATCAATGCTGTGTACTGCTAGTCTGTAATGGGATG	1397
Db	1339	TCGAATCAATTCGTTTCGAGAGTGATCAATTAACGGGGTTACTGCTAGTCTGAGTGGGATG	1398
Qy	1398	TAGATAAGCAATCTTCAACCCCTTGTCAAATGACTTTCTTCTTGACAGCTAAGCTGGCTG	1457
Db	1399	TCGATAAATCATTTATACAGTCCCTTGTCATGAGCTTCTTCTCTACACACCAAGCTTCGG	1458
Qy	1458	CGCTTCAAGACGATATAGTAATAGGAAGTTTCGGTGTGGATPAGACCACTTCTCGAAC	1517
Db	1459	TTCTGAAAGACGATCTTTTGTATAGCAAGTTTGCATCTGGACCAAAAACCTGCTCTCAAC	1518
Qy	1518	TTATTTGGGATAGGTGGGCAATTTTTTGGAAACGTTTTTCCCACTATCAAAGAGAGAT	1577
Db	1519	ATGTGTGGGATAGAAITTCCTAGCTTTCCGCAATGCTTTTCCCAATCGATCAAGGAAGAT	1578
Qy	1578	TGGTGACGAGCAAAATCTCGATGTAAGTGAGATGCTCTGGAAGATCAAGATCCAGATC	1637
Db	1579	TGATAAACCCGAAACTGATCAAAATATCGGAGATGCGTTAGAGATCAGGGTGCCTGATC	1638
Qy	1638	TGATGTGCACATGGAAGACAGGTTTCGTAGCTGTAATACACCAAGCTCTGAGGAGTTACCGC	1697

[illegible]

Db 2719 GCACCGTCGATTCATTTTGTGATGAATACCGGAAAGGGGACACGCTGTCTCAGTTCAAAGAT 2778  
Qy 2763 TTTTATTTGATGAAGGGTTGATGCTGCACACCGTTGTGTAACTTCTGTTGCTTATCT 2822  
Db 2779 TGTTCATAGACGAGGTTTGTGCTGCTACATCTGTTGTGTGAATTTTGGTTGAAATGT 2838  
Qy 2823 CTGGTTGCGACATCGCATACATTTACGGAGATACACAGCAGATTCCTTTTCATTAACAGAG 2882  
Db 2839 CTCTGCGCATATTCGATATGTTTATGGAGACACCCAAACAGATTCGTTACATCAACAGAG 2898  
Qy 2883 TTCAGAAATTTCCGGTATCCAAACATTTTGAAGAGCTGCAAGTGATGATGATGA 2942  
Db 2899 TAACTGGTTTCCCGTACCCCTGCACACTTTTGCAAAATTTGAGGTGACGAAAGTCGAAACAA 2958  
Qy 2943 GGAGGACCACTGAGATGCCAGGTGATGTCAAATTTTCTCAATCGAAGTACGAG 3002  
Db 2959 GAAGAACTACTCTTGTGTGCGGCTGATGCACACTTCTCTAAATCAAGGTATGAAG 3018  
Qy 3003 GAGCGGTGACAACTTCAACTGTACAAACGATCGGTCTCATCTGAGATGATPAGCGGTA 3062  
Db 3019 GACAGTAAATGTGCACGCTTCTCTGAAAGAAATCAGTTTCCAGGAAATGGTTAGTGGG 3078  
Qy 3063 AGGAGTACTAAACAGTGTTCCTCAACCACTAAAGGGAATTTGTAACCTTCACTCAGG 3122  
Db 3079 CTGCGTCTATCAATCCTGTGTCCAGCCGCTTAAAGGGAATTTTGAATTTTCAACAGT 3138  
Qy 3123 CTGATAAATTTGAGTTAGAGGAGAGGGCTATAAGAAATGTGAACACCGTTCAATGAGATCC 3182  
Db 3139 CTGACAGGAGGCCCTTCTCTCAAGGGCTACGAGATGTCATGATGATGAGGTAC 3198  
Qy 3183 AAGGAGAACTTTGAAGATGTGCGGTGAGATGACGGCAACTCCACTGACTCTGA 3242  
Db 3199 AAGGTGAGACTTATGACAGCGTATCGTTAGTTCGACTAACCACTACGCTGTATCTATCA 3258  
Qy 3243 TTTTCAAGTCTTCCCGCATGTTCTAGTCGCTCTGACTAGACACAAAGAGCTTCAAT 3302  
Db 3259 TCGCAAGAGACAGTCCGATGTTCTGCTGTTGTCAGGACACACAAAATCCCTAAAGT 3318  
Qy 3303 ATTACACCGTAGTTAGATCCTTTAGTACAGATAATTAGTGTGCTTCTTTAAGCT 3362  
Db 3319 ACTACACCGTTGTGATGATCCTTTAGTTAGTATCATTAGAGATTTAGAACGGGTAGTA 3378  
Qy 3363 CTTTCTTTTGAATATATGTGTAGAGACGAGTAGTAGATAGCAATTAACAGATGGATG 3422  
Db 3379 GTTACTTATTAGACATGTACAAAGTAGATGACAGGTACTCAATAGCAATTAACAGGTGACT 3438  
Qy 3423 CAGTCTTCAAAGTCAATATCTTTTGTGGCAACCTAAATCAGGAGACTTCCAGATC 3482  
Db 3439 CTGTTTAAATAATTTCAATCTTTTGTAGCAGCTCCAAAGACTGGAGATATATCTGATA 3498  
Qy 3483 TACAGTTCTATPACGATGTATGCTCCCTCGTAAATAGTACTATATTAACCAAGTATGATG 3542  
Db 3499 TGCAATTTTACTATGATAAGTGTCTTCTGGGAACAGCACGTTGTTGAACAACTACGACG 3558  
Qy 3543 CTGTTACCATGAGGTTAGTGATATAGTCTTAATGTGAAGGATGTGTTCTTGATTTTT 3602  
Db 3559 CTGTTTACCATGAATTTGACTGACATTTCTCTGAATGTCAAAGATTTGCAATATAGATATGT 3618  
Qy 3603 CCAAAAGTATTCGGATGCTCAAAGGAGTGAAACCAATGTCTAGAGCCAGTTTGGTACCG 3662  
Db 3619 CTAGTCTGTAGTCTCCGAAGATGTCAACCAACTTAAATCCGATGTGCAAGCG 3678  
Qy 3663 CGCGGAACCGCAAGGCTCGAGACTTCTCGAAATCTGGTTGCAATGATTTAAAGAA 3722  
Db 3679 CGGCAGAAATGCCTCGCCAGACTGAGCTGTTTGGAAATCTAGTTGCGATGATTTAAAGAA 3738  
Qy 3723 ATTTCAACGACCACTGACCGGAGCATGATGACATTTGAGAGCACCGCATCTGTTGATG 3782  
Db 3739 ATTTTAAATCCACAGATGTGCGGAGTAGTTGATTTGAAATTAACATCTTTAGTGG 3798  
Qy 3783 TAGATAAGTTTTTGTAGTATTTTATTAAGAAAGAAATTAACAAAAATATTTGCTG 3842  
Db 3799 TAGATAAGTTTTTGTAGTATTTTACTTAGGAAAGAAAGAAACCAACAAAAATTTTT 3858

Qy 3843 GAGTGATCACGAAGGATTTCAATGATGAGATCGTTGCGAAAACAGGAAAGAGTACTATTGG 3902  
Db 3859 CACTGTTTAGTAGAGATCTCTCAATAGGTGATAGCAAGCAAGAACAGTCAAAATG 3918  
Qy 3903 ACGACTTGGCTAACTTACAAATTTTACAGATCTGCGGGCCATCGATCAGTACAGACACATGA 3962  
Db 3919 GTCAGTTGGCCGATTTTGTATTTTGTGGATCTTCCAGCCGTTGATCAGTACAGGCATATGA 3978  
Qy 3963 TCARGGCTCAACCAAAACAGAAAATTTGGACCTTTCAATTTCAAGTAAGTAACCTGCTCTGC 4022  
Db 3979 TTAAGGCCAACCCGAAGCAGAAATCTGGATCTGTCAATTTCAAGACAGATATCCAGCGTTGC 4038  
Qy 4023 AAACAATTTGCTTACCATTTCGAAGCAGATCAACGGTATTTTGG--CCGGTTTCTCAGAG 4079  
Db 4039 AAACGATTTGTATCTCAATCAAGAAATAATCAACGCAATATTTGGTCTCTTTTTCAGTGAGC 4098  
Qy 4080 TTACAAGGTTGCTGCTCGAGGCAATTTGATTTCTAAGAAAGTTTCTTTTCTTACTAGGAAAA 4139  
Db 4099 TTACAAGGCAATTTACTTGCACAGTATTTGACTCAAGCAGATTTCTTGTCTTTTACGAGAAAGA 4158  
Qy 4140 CTCCAGAACACATTCAGAAATTTTCTCGGATCTGACTCCGACGTTCCCTATGATGTGT 4199  
Db 4159 CACCGGCTCAGATCGAAGATTTCTTCGGAGATCTAGACAGTCATGTCCCAATGACGCTAC 4218  
Qy 4200 TAGAACTGGATATTTCTAAGTATGATAAGTCACAGAACGAGTTTTCATTGTCTGTAGAGT 4259  
Db 4219 TTGAGTTGGATGTTTCGAAGTATGATAAGTCTCAAAACGAGTTTTCATTGTCTGTGAGT 4278  
Qy 4260 ATGAAATATGAAAGATTTGGGTCTCAATAGTATTTTGGCCGAAGTGTGGAAAAACAAGGC 4319  
Db 4279 ACGAAATCTGGAGGAGACTGGGTCTGGAGGATTTCTTTGGCAGAAAGTGTGGAAAAACAAGGC 4338  
Qy 4320 ACAGGAAACAACTTTGAAAGATTTACATTCGTGGAATCAAGACATGCTGTGCTATCAAA 4379  
Db 4339 ATAGAAAACCACTCTGAAAGATTTACATCTGTGTATAAAAAACGTGTTATGTTGATCCAGA 4398  
Qy 4380 GGAAGGCGGTGATGTGACTACTTTTCATCGCAATACTGTTATATAATAGCAGCTTGTCTGG 4439  
Db 4399 GAAAGAGTGGTGTATGTTTACAACTTTTATCGGTATATACCGTCATCATTTCTGCTGTCTAG 4458  
Qy 4440 GTTCAATTTTACCAGTGAAGGATCAATAAGGTGCTTTTGTGGAGACGATTTCCGTTT 4499  
Db 4459 CATCAATCTCCCGATGGAATAATTTGATAAAGGAGGCTTCTGCGGAGATGACAGTTTGT 4518  
Qy 4500 TGTATTTTCCAAAGGGTTTGGATTTTCCCTGACATTCAGTCAATGCTTAATCTCATGTGA 4559  
Db 4519 TGTAATTTTCTAAGGGTTGTGATATCCCGATATACAAACAGCTGCTTAATCTAATGTGGA 4578  
Qy 4560 ATTTTGAAGGCCAACTGTATAGAAAGAGGTACCGTTACTTTTGTGGTAGATACATCATAC 4619  
Db 4579 ATTTTGAAGGCCAACTGTTCAGAAAGCAATATGSGTACTTCTGCGGGAGGTACGTGATTC 4638  
Qy 4620 ACCATGATAGAGGAGCAATATGATGATTAATGATCTTTGAAAGTGTGATCTCCAAACTTGGGG 4679  
Db 4639 ATCAACGATAGAGGTTGATAGTATATACGACCTTTTGAAGCTGATTTTCCGAACTTGGTG 4698  
Qy 4680 CAAACATATCAAGGATTTATGATCAGTACAGGATTAAGGGTCTTTTGTGCGATGTTG 4739  
Db 4699 CTAAACATATCAAGGATTTGGATCATTTGGAGGAGTTTCAAGATCCCTCTGTGATGTTG 4758  
Qy 4740 CTTGTTCCGCTCGGAACTGGTGTCTTAGGCTTTCCGACAGCTGAACGACGATATCAAGGAGG 4799  
Db 4759 CTGAGTCGTT--GAACAAATTGCGCGTATTACACAAATTTGGACGCGCTGTGGGAGG 4815  
Qy 4800 TTTCAAAAACCGGATTTGATGTTGCTGTTGTTTAAATTTGTTTAAATAAATTTTGTGTG 4859  
Db 4816 TTTCAAAAACCGGCCCCACCTGGTTCGTTGTTTATAAGAGTTTAGTTAAGTATTTTGTGAG 4875  
Qy 4860 ATAAATTTTATTTAGAACTTTTGTTTTAAATGGCTGTAGTCTCAGAGATACGTCTCAAA 4919  
Db 4876 ATAAAGTTTTTGTAGAGTTTATTTCTTCTAGTCTAGTTTGTAAAGGTAGGTTAAT 4935

```
QY 4920 ATTACGAGTTCAATGATCTTTGAAACAGGATGAGATACCTCCGGCAATTCATGACTAAG 4979
DB 4936 ATTAATGAGTTTATCGATCTGTGCAAAAGTCTGAGAAACTTCTCCGCTCGATGTTCAACGCT 4995
QY 4980 GTCAAGAGTGTAGAAATCACTGTGCAAGATTAATGGCTGTGTAAGAAATGATAGTCTT 5039
DB 4996 GTAAAGAGTGTATGGTTTCAAGGTTGATAGATTAATGGTCCATGAANAATGAATCATTTG 5055
QY 5040 TCTGATGATGATTTACTTTAAAGGTGTTAAAGTTAGTTAGTAAAGAAAGGTATGTTGCTTAGCT 5099
DB 5056 TCTGAAGTAAATCTCTTAAAGGTGTAATACTTATAGAAGGTGGGTATGTTGCTTAGTGC 5115
QY 5100 GATTGGTAGTGTCTGGGAGTGAATCTCCGATTAATCTCCGCTGGTGGTGTGAGTGT 5159
DB 5116 GGTCTGTGTGTCCGGTGGTGAATTTTACAGATAATTCGCGTGGTGGTGTGAGTGT 5175
QY 5160 TGTATTGTAGATATAGAGAAATGAAAGAGTAGTAAGGAAGCAAGCTGGGTGCGTATCACGCC 5219
DB 5176 TGCATGTTGACAGAGATATGGAAGAGCGGACGAGCCACACTGGGTGATATTAACACT 5235
QY 5220 CTTGCTGCAAAAAAGAAATTTTCTTTTAAAGCTAATCCCTTAATTAATCAATAACATCCGAG 5279
DB 5236 CTGCTGCTGCAAAAAAGCGTTTCAGTTTAAAGTGTGCCAAATTTACGGTATTTACAAACAAG 5295
QY 5280 GATGCTGAGAGACCCCGTGGCAAGTGTAGTGAATATCAAGAGAGTGGCTATGGAAGAA 5339
DB 5296 GATGCAAAAAAGAACATATGCGAGGCTCTTAGTAATATTAATAATTTAAATAATGTAAGATGCG 5355
QY 5340 GGATACTGCTCTTATCTTTGAGTTCGTTTCAATTTGTTAGTACATATAAAAAATTAAGTGA 5399
DB 5356 GGCTACTGCCCTTTGCTCAATTAGAAATTTGTTGTTGTTGTTTATTAATAATAATAATA 5415
QY 5400 AGAAAGGTTTGGAGGAAACGTAATTTTGAAGTGTGACAGCGGCTCGCCAAATTTGAACACT 5459
DB 5416 AAATTTGGGTTTGGGAGAAAGTAACGAGTGTGAACGATGAGGAGCCATGAACTTTTCG 5475
QY 5460 GAAAAGTTTGTAGGAGTTCGTGATGAAGTACCAATGGCTGTGAAACTCGAAAAGTTT 5519
DB 5476 GAAGAAGTTTGTAGTGTTCATGAGAAATGTTCCAAATGTCGGTTAGACTCGCAAAGTTT 5535
QY 5520 CCGGAAAAACAAAAAGAAATGTTAGTAAATATGTTAATAAT----- 5561
DB 5536 CGAACCAATCTCTCAAAAGAGTTCGGAATAATAATAATAATTTAGTAAAGGGGGTTCA 5595
QY 5562 ---AAGAAATAAATAACAGTGGTAAGAGGGTTTTTAAATTTGAGGAAATTTGAGGATAAT 5618
DB 5596 GGCGGAAGGCTTAAACCAAAAAGTTTGTATGAAGTTGAAAAGAGTTTGTATAATTTGATT 5655
QY 5619 GTAAGTATGATGAGTCTATCGGTCATCGAGTAGCTTTTAATCAATATGCTTATACAA 5678
DB 5656 GAAGATGAAGCCGAGACGTCGGTCGCGGATTTCTGATTCGTATTAATAATGTTCTTACTCAA 5715
QY 5679 TCAACTCTCCGAGCCAAATTTGTTTACTTATCTTCGCTTACGAGATCCTGTGAGCTGA 5738
DB 5716 TCACCTCTCCATCGCAATTTGTGTTTGTATCTGTATGGGCTGACCCATAGAAATGT 5775
QY 5739 TCAATCTGTGTACAAATGCAATTTGGGTACCACTTTCAAAACGCAACAACTAGGCAACAG 5798
DB 5776 TAAACGTTTGTACAAATTCGTTAGTAAACAGTTTCAAAACAGCAAGCAAGCACTACTG 5835
QY 5799 TCCAAACAGCAATTTGCGGATCGCTGGAAACCTGTGCTAGTATGACAGTGAATTTCTGT 5858
DB 5836 TTCAACAGCAGTTTCAGCGAGGTGTGGAAACCTTTCCCTCAGACGACCCGTCAGATTTCTGT 5895
QY 5859 CATCGATTTCTATGTATGATATATTAATTCGACGCTTGATCGCTTGATCAAGCGGTTAT 5918
DB 5896 GCGATGTTTAAAGGTGACAGGTACAAATGAGTTTATGATCCTCTAATTAATCTGCGTTC 5955
QY 5919 TAAATAGCTTTGATACTAGAAATAGAAATAATAGAGGTTGATAATCAACCCGACCGGATA 5978
DB 5956 TGGGGCTTTTGTACTAGTAAGTAAGTAATTCGAGTAGTAATAACAGCAGAGTCCGACAA 6015
QY 5979 CTACTGAAATCGTTAAACCGGACTCAGAGGGTAGAGATGCTACTGTACTATAGGGGCTT 6038
```

```
DB 6016 CAGCTGAACGTTAGATGCTACCCGAGGGTAGACGCTACGTTGCAATTCGGTCTG 6075
QY 6039 CAATCAATAATTTTGGCTAAATGAATGGTTTCGTGGAACCTGGCATGTTTCAATCAAGCAGGCT 6098
DB 6076 CTATAAATAATTTAGTTAAATGAACACTAGTAAGAGGTACTGGACTGTACAAATCAGAATACTT 6135
QY 6099 TTGAGACTGCTAGTGGACTTGTCTGGACCACAACTCCGGCTACTTTAGCTATTTGTTGTGAG 6158
DB 6136 TTGAAGATGATCTCTGGGTTGGTCTGGACCTCTGCACCTGCATCTTTAAATGATAGGTGCT 6195
QY 6159 ATTTCTTAAATAAAGTCGCTGAAGACTTAAATTTCAAGGTGGCTGATACCAAAATCAGC 6218
DB 6196 GAAATATAAA-----GTTTGTGTTTCTAAAAACACACGTTGATACGATTAACGTAC 6247
QY 6219 AGTGGTTGTTTCCGCTCCACTTAAATATAACGATTTGTTCATATCTGGATCCAAACAGTTAAACCA 6278
DB 6248 AGTGTGTTTTCCTCCACTTAAATCGAAGGTAGTGTCTTTGGAGCGCGGAGTAAACATA 6307
QY 6279 TGTGATGGTGTATATCTGTGTGATGCGGTAAACATCGGAGAGGTTTGAATCCTTCCCTAA 6338
DB 6308 TATGTTTCAATATATGTCGTAGGACGCTAAAAAAGGAGGATTCGAATTTCCCCCGNA 6367
QY 6339 CCGCGGCTAGCGGCCA 6355
DB 6368 CCCCCGTTGGGCCCA 6384

RESULT 4
ABA04211
ID ABA04211 standard; DNA; 6384 BP.
XX
AC ABA04211;
XX
DT 01-MAR-2002 (first entry)
XX
DE Tomato mosaic virus related DNA sequence.
XX
KW Tomato mosaic virus; vaccine K; low-virulent vaccine K genome;
point mutation; replicase; motor protein; plant virus genome;
genetic engineering; viral disease; plant; ds.
XX
OS Tomato mosaic virus.
XX
PN CN1306090-A.
XX
PD 01-AUG-2001.
XX
PF 14-JAN-2000; 2000CN-00100211.
XX
PR 14-JAN-2000; 2000CN-00100211.
XX
(MICR-) INST MICROBIOLOGY CHINESE ACAD SCI.
XX
PI Qiu B, Yang G, Tian B;
XX
WP1; 2002-000227/01.
XX
PT Low-virulent vaccine K genom sequence of tomato mosaic.
XX
PS Example 4; Fig 1 (disclosure); 22pp; Chinese.
XX
CC The present invention describes the genome separated and cloned from
tomato mosaic virus, which is modified into low-virulent vaccine K genome
by point mutation. The nucleotide sequence of its coded replicase and
motor protein has the mutation, so it can be used to modify the plant
virus genome and prepare genetically engineered vaccine for effectively
preventing and treating the viral diseases of plants. The present
invention represents a DNA sequence which is given in an example from the
present invention
XX
SQ Sequence 6384 BP; 1906 A; 1185 C; 1478 G; 1815 T; 0 U; 0 Other;
```





Db 2119 CAGCTAGTTCCTTAATTCATTAAGCAGATGTGTTTCGATCGGTACACGCGGCCCTCTTAAAG 2178  
Qy 2172 TTCAACAAATGAAGAACTATGTGGACAGTTCGTTGGCAGCTTCGTTGTCGCCCACTGTATCAA 2231  
Db 2179 TTCAACAAATGAAGAACTTTATAGACAGCCTGGTAGCCTCGCTCTCTGCTGGGTGCGA 2238  
Qy 2232 ATCTATGCAAGTCACTAAAGATGAAGTCGGGTATGATCTTGATTCAGGGAGAAAGTTG 2291  
Db 2239 ATCTAGTGAAGATCCTTAAAGATACAGCCGGATTTGACCTTGAACCTCGTCAAAAGTTGCG 2298  
Qy 2292 GTGTTTGGGATGTCACTTTTGAAGAGTGGCTCCTCAAACTCGCGCCCAAGAGGTCAATTCAT 2351  
Db 2299 GAGTTCTGGATGTGCTTCGAAAGGTGGCTAGTTAAACCAATCCGCAAGAGCAATGCAT 2358  
Qy 2352 GGGGAGTTGTCTCGATTAACAAGGGGAAATGTTTTACTGCACTTCTATCTTATGAAGGAG 2411  
Db 2359 GGGGGTTGTTGAGACTCATCGAGGAAATATCGCGTCGCATTACTGCGAGCACGATGAAT 2418  
Qy 2412 ATAGAAATGGTCACTGAGAGCACTGGAGGAGGTGGCTGTATCATCTCATCAATGGTAT 2471  
Db 2419 TTGGCATTTATACGTGCGATAACTGGCGACGGGTGGCTGTGAGTTCTGAGTTCGGTAGTAT 2478  
Qy 2472 ATTCTGATATTGCAAAAGCTCCAAATCTGAGGAAACAAATGAGAGCGGTGAACCCCAAG 2531  
Db 2479 ATTCTGATATGCTAAACTCAGGACTCTGAGAAGATTGCTCAAGATGGAGAACACACAG 2538  
Qy 2532 AACCTACTGCAAGATGTGACTTTGTGGATGGGGTGCCTGTTGTGGAAAGTACAAAGGAG 2591  
Db 2539 TTAGTTCAAGCAAGGTGGTTTTGTTGGATGGCTTCCAGGGTGGGGAGAACAAAGGAA 2598  
Qy 2592 ATTTTGAAGATTGATCTTGATGAGGATTTGATCTTGGTTCTGGGAAAAACAAGTGTCTG 2651  
Db 2599 TTCTTTCGAGAGTTAAATTCGAAGAGAACCTTAATTTCTTGTCCCTGGTGTCAAGTGTCCG 2658  
Qy 2652 CTATGATCAGAGAGGCTTAATTCATCTGACTGATGAAGACCAATGACATGTA 2711  
Db 2659 AGATGATCAGATGAAGACTAATGGCTCGGCAATATAGTGGCTACAAGGATTAATGTGC 2718  
Qy 2712 GAACGGTAGATTCACTTCTAATGCAATCCAAACCGCGATCAC-----ACAAGAGGC 2762  
Db 2719 GCACGTCGATTCATTTTTTGTATGAATACGGGAAAAAGGCGCGCTGTCAAGTTCAAAAGAT 2778  
Qy 2763 TTTTATTTGATGAAGGTTGATGTGACACAGGTTGTGTTAACTTCCTGGTGTCTATCT 2822  
Db 2779 TGTTTCATAGACGAAGGTTGATGTGCACTGCTGGTTGTGTGAATTTTTTGGTGTGAATGT 2838  
Qy 2823 CTGGTTGGACATCGCATACATTTACGGAGATACACACAGATTCCTTTCAATTAACAGAG 2882  
Db 2839 CTCTGGCGATTTGCATATGTTTATGAGACACCCAAACAGATTCGGTACATCAACAGAG 2898  
Qy 2883 TTCAGAATTTCCCGTATCCCAACATTTTGAAGAGCTGCAAGTGGATGAAGTTGAGATGA 2942  
Db 2899 TAACTGGTTTTCCCGTACCCTGCACACTTTGCAAAATTTGGAGTGCAGCAAGTCGAAACAA 2958  
Qy 2943 GGAGACCACTGAGATGCCAGGTGATGGAATTTTTTTCGTCATCAATCGAAGTACGAAG 3002  
Db 2959 GAAGAACTACTCTTCGTTGTCCGGCTGATGTCAACACTTCTTAAATCAAAAGGTATGAAG 3018  
Qy 3003 GAGCGGTGCACAACCACTTCAACTGTACAAACGATCGTCTCATCTGAGATGATAGCGGTA 3062  
Db 3019 GACAGTAATGTGCACGTCTTCTGAAAGAAATCAGTTTCCAGGAATTTGTTAGTGGGG 3078  
Qy 3063 AGGGAGTACTAAACAGTGTTCCAAACCACTAAAGGGGAAATTTGAATTTCACTCAGG 3122  
Db 3079 CTGGCTCTATCAATCCTGTGTCAGCGCTTAAGGGAAAAATTTTGACTTTCACACAGT 3138  
Qy 3123 CTGATAAATTTGAGTTAGGAGAGAGGCGTATAAGAAATGTGAACACCGTTCAATGAGATCC 3182  
Db 3139 CTGACAAGAGGCCCTTCTCTCAGGGGCTACGCAGATGTCCATCTGTACATGAGGTAC 3198  
Qy 3183 AAGGAGAACCTTTGAAGATGTGCTGCTCAGATTGACGGCACTCCACTGACTCTGA 3242  
Db 3199 AAGGTGAGACTTATGACAGCGTATCGTTAGTTTCGACTAACCTTAACCTTACGCCCTGTATCTATCA 3258

Qy 3243 TTTCCAAAGTCTTCCCCGCAATGTTCTAGTCGCTCTGACTAGACACACAAAGAGCTTCAAAT 3302  
Db 3259 TCGCAAGAGACAGTCCGCAATGTTCTGTGCTCTGTTGTCAAGACACACAAAAATCCCTAAAGT 3318  
Qy 3303 ATTACACCGTAGTCTAGATCCCTTTAGTACAGATAAATTTAGTGAATTTGCTCTCTTTAAAGCT 3362  
Db 3319 ACTACACCGTTGTATGGATCCCTTTAGTTAGTATCATTTAGAGATTTAGAACGGGTAGTA 3378  
Qy 3363 CCTTCCCTTTTAGAANAATGTATATGTAGAGCAGGTAGTAGTAGCAATTTACAGATGGATG 3422  
Db 3379 GTTACTTTATTAGACATGTACAAAGTAGATGACAGTACTCAATAGCAATTTACAGGTCGACT 3438  
Qy 3423 CAGTGTTCAAAGGTCATAATCTCTTTGTGGCAACACCTAAATCAGGAGACTTTCCAGATC 3482  
Db 3439 CTGTGTTTAAAAATTTCAATCTTTTGTAGCAGCTCCAAAGACTGGAGATATATCTGATA 3498  
Qy 3483 TACAGTTCTATTTACGATGTATGCTCCCTGCTGAATAGTACTATCTTAAACAAGTATGATG 3542  
Db 3499 TGCAATTTTACTATGATAAGTGTCTTCTCTGGAAACAGCACGTTGTTGAAACAACTACGACG 3558  
Qy 3543 CTGTTACCATGAGGTTTACGTGATAATAGTCTTAAATGTGAAGGATTTGTTCTTGTATTTT 3602  
Db 3559 CTGTTACCATGAATTTGACTGACATTTCTCTGAAATGTCAAAGATTTGCATATTTAGATATGT 3618  
Qy 3603 CCAAAAGTATTTCCGATGCGCAAAAGGAGTGAAACCAATGCTTAGAGCCAGTTTTCGTACCG 3662  
Db 3619 CTAAGTCTGTAGCTGCTCCGAAGATGTCAAACCAACTTTAATACCGATGGTACGAACGG 3678  
Qy 3663 CGCGGAAACCCCAAGGGCTGCAGGACTACTCGAAATCTGGTTGCAATGATTTAAAGAA 3722  
Db 3679 CGGCAGAAATGCCTCGCCAGACTGGACTGTTGGAAATCTAGTTGCGGATGATTTAAAGAA 3738  
Qy 3723 ATTTCAACGCAACACAGACCTGACGGGGAGATTGACATTTGAGAGCACCCGATCTGTTGTAG 3782  
Db 3739 ATTTTAAATTCACAGAGTTGTCCGAGTAGTTGATATTGAAATACTGCACTCTTTAGTGG 3798  
Qy 3783 TAGATAAGTTTTTGTAGTAGCTATTTTATTAAGAAAGAAAAATACACAAAAAATATTGCTG 3842  
Db 3799 TAGATAAGTTTTTGTAGTAGTTATTTACTTTAAGGAAAAAAGAAAAACCAACAAAAATTTT 3858  
Qy 3843 GAGTGTAGCAACGAAGGNTCAATGATGATGATGTTGGAAAAACAGGAACGAAGTACTATTGG 3902  
Db 3859 CACTGTTTAGTAGAGAGTCTCTCAATAGTGGATAGCAAGCAAGACAGTCACAATTTG 3918  
Qy 3903 ACGACTTGGCTAACTCAAAATTTTACAGATCTGCGGGCCATCGATCAGTACAGCAATGA 3962  
Db 3919 GTCAGATGGCCGATTTTGAATTTTGTGGATCTTCCAGCGTTGATCAGTACAGGCATATGA 3978  
Qy 3963 TCAAGGCTCAACCAAAACAGAAATTTGGACCTTTCAATTTCAAGATGAATACCTGTCTGTC 4022  
Db 3979 TTAAAGCGCAACCGAAGCAGAACTGGATCTGTCAATTTCAAGCAAGATATCCAGCGTTGC 4038  
Qy 4023 AAACAAATTTGCTCACTCCATTTTGAAGCAGATCAACGGTATTTTGG---CCGGTTTTCAGAGC 4079  
Db 4039 AAACGATTTGTATCATTTCAAGAAATCAAGCAATATTTGGTCTCTTTCAGTGAGC 4098  
Qy 4080 TTACAAGGTTCTCTGCTCGAGGCAATTTGATTTCTAAGAAGTTTCTTTTCTTACTAGGAANA 4139  
Db 4099 TTACAAGCAATTTACTTGACAGATTTGACTCAAGCAGATTTCTTGTCTTTACGAGAAAGA 4158  
Qy 4140 CTCAGAAACAGATTCAGAAATTTTTTCTCGGATCTCGATCTCGACGCTTCATGATGATGT 4199  
Db 4159 CACCGGCTCAGATCGAAGATTTCTACGGAGATCTAGACAATCATGTCCTCAATGGACGTAC 4218  
Qy 4200 TAGAACTGGATATTTCTAAGTATGATAGTACAGACAGTTCATTTGCTGCTGTAGAGT 4259  
Db 4219 TTGAGTTGGATGTTTTCGAAGTATGATAAGTCTCAAAACGAGTTTCTTTGCTGTGTAGT 4278  
Qy 4260 ATGAAATATGAAAGAGATTGGGTCTCAATGAGTTTTTGGCCGGAAGTGTGGAAACAAGGGC 4319  
Db 4279 ACGAAATCTGAGGAGACTGGGTCTGGAGGATTTCTTGGCAGAGTGTGGAAACAAGGGC 4338

QY 4320 ACAGGAAACAACTTTGAAGGATTAACATCTGCTGGAAATCAAGACATGTCGTGGATCAAA 4379  
Db 4339 ATAGAAAAACCACTCTGGAAGATTAACATCTGGTATATAAAACGCTTTATGGTACCAGA 4398  
QY 4380 GGAAGAAGCGGTGATGTGACTCTTTCATCGCAATATCTGTATATATAGCAGCTTCCTGG 4439  
Db 4399 GAAAGAGTGGTGAATGTTACACTTTATCCGGTAATACCGTCATCATGCTTCGTGCTAG 4458  
QY 4440 GTTCAATGTTACCGATGGAAAAAGGTCATAAAAGGTCCTTTTGTGGAGACGATTCGGTTT 4499  
Db 4459 CATCAATGCTCCCGATGGAAAAATTCATAAAGAGGCTTCTCGGAGATGACAGTTGCT 4518  
QY 4500 TGTATTTCCAAAGGTTTGGATTTCCCTGACATTCAGTCATGTGCTAACTCATGTGA 4559  
Db 4519 TGTACTTTCTTAAGGTTGTGAGTATCCCGATATACAAAGAGCTGTAATCTAAATGTGA 4578  
QY 4560 ATTTTGGAGCCAACTGTATAGAAAGAGGTACGGTACTTTTGTGGTAGATACATCATAC 4619  
Db 4579 ATTTTGGAGCCAACTGTTCAAGAGCAATATGGGTACTTTCGCGGAGGTACGTGATTC 4638  
QY 4620 ACCATGATAAGGAGCAATGATGTATATGATCCTTTGAAAGTTGATCTCCAAACTTTGGG 4679  
Db 4639 ATCAGATAGAGTTGCTAGTATACTACGACCTTTGAGCTGATTCGAACTTTGGTG 4698  
QY 4680 CAAAACATATCAAGGATTAATGATCACTTAGAAGAGTTAAGGTTGCTTTGTGCGAATMG 4739  
Db 4699 CTAAACACATCAAGGATTTGGAGTCAATTTGGAGAGTTTCAGAAAGATCCCTCTGTGATGTG 4758  
QY 4740 CTGTGTTGCTCGAAACTGGTCTTAGGCTTTCCGAGCTGAACCGCATCAAGGAGG 4799  
Db 4759 CTGAGTCTGTTT---GAACTATTCGGGTATTACACAAATGAGCAGCTGTTGGGGAGG 4815  
QY 4800 TTCTATAAAACCGCGATGATGTTGCTGTTGCTTTTAAATTTGTTTAAACAAATTTTGTGTG 4859  
Db 4816 TTCTATAAAACCGCCACCTGTTGCTGTTGTTTATAAGAGTTTGTAAATTTGTCAG 4875  
QY 4860 ATAAATTTTATTTAGAACTTTGTTTAAATGGCTGTTAGTCTCAGAGATACCTGTCAA 4919  
Db 4876 ATAAAGTTTGTGTTAGAAATTTATTTCTTGATGGCTCTAGTTGTTTAAAGGTAAGGTAAT 4935  
QY 4920 ATTACGAGTTCATTTGATCTTTTCGAAACAGAGTATGATCTTCCGCGATTTCATGACTAAG 4979  
Db 4936 ATTAATGAGTTTATCGAICTGTCGAAAGCTGAGAACTTCTCCCTCGATGTTACGGCT 4995  
QY 4980 GTCAAGAGTGTAGAAATTCGACTGTGGAACAGATTTATGGCTGTTTAAAGATGATAGTCTT 5039  
Db 4996 GTAAAGAGTGTATGTTTCAAGGTTGATAAGATTATGGTCCATGATAAATGAATCATTG 5055  
QY 5040 TCTGATGTAGATTTACTTTAAAGTGTTAAGTTAGTTAAGAAAGGTATGTGTCTAGCT 5099  
Db 5056 TCTGAAGTAAATCTCTTAAAGGTGTAACACTTATAGAAGGTGGGTATGTTTGTCTTAGTC 5115  
QY 5100 GATTTGGTAGTCTCTGGGAGTGAATCTCCCGATAACTCGCGGTGGTGTGCTAGTGT 5159  
Db 5116 GGTCTGTTGTCGGTGTAGTGAATTTACAGATAATGTCGGTGGTGGTGTGAGTGTG 5175  
QY 5160 TGTATTTAGATAAGAGAAATGAAGAAGGATGAAGGAAGCAACGCTGGGTGCGGTATCACGCC 5219  
Db 5176 TGCATGTTGACAAGAGAAATGAAGAAGCGGAGCGAAGCCACACTGGGTCTATATTACACT 5235  
QY 5220 CTTGCTTCGAAAAAGAAATTTTCTTTTAAAGCTAATCCCTAATATTCAATTAACATCCGAG 5279  
Db 5236 GCTGCTGTAAGAACGGTTTTCAGTTTAAAGTGGTCCCAAAATACCGTATTACACAAAG 5295  
QY 5280 GATGCTGAGAAGCACCGCTGGCAAGTGTAGTGAATATCAAGGAGTGGCTATGGAAGA 5339  
Db 5296 GATGAGAAAGAAACATATGCGAGTCTTAGTAATATATAAATGTAATAATGATGTCG 5355  
QY 5340 GGATCTGTCTTTATCTTTGGAGTTCGTTTCAATTTGTGTAGTACATAAAAAATAATGTA 5399  
Db 5356 GGCTACTGCCCTTTGTCATTAAGAAATTTGCTGTGTGTATTTGTTTATAAAAAATAATA 5415  
QY 5400 AGAAAAGGTTTCAGGGAACGATTTTGTAGTGTGACAGACGGCTCCGCAATTTGAATCACT 5459

Db 5416 AAATTTGGTGTGGAGGAGAAAGTAACGAGTGTGAACGATGGAGGACCCATGGAATTTTCG 5475  
QY 5460 GAAAAGGTTGTTGAGGAGTTCGTGGATCAAGTACCAATGGCTGTGAAATCGAAAGGTT 5519  
Db 5476 GAAGAAGTGTGATGAGTTCATGGAGATGTTCCAAATGTGCGTTAGACTCGCAAGTTT 5535  
QY 5520 CCGGAAAAACAAAAAGAAATGTTAGTAAATGTTTAAATAAT----- 5561  
Db 5536 CGAACCAATCTCAAAAGAGGTCGGAATAAATAATAATAATTTAGGTAAAGGGCGTTCA 5595  
QY 5562 -----AGAAAAATAAATAACAGTGGTAAGAGGTTTTAAATTTGAGAGAAATGAGGAT 5615  
Db 5596 GCGGAGAGGCTTAAACCAAAAAAGTTTTCATGAAGTTTAAAAAGAGCTTTAAATAATTTGAT 5655  
QY 5616 AATGTAAGTGTGATGACGAGTCTATCCGTCATCGAGTACGTTTAAATCAATATGCTTATA 5675  
Db 5656 AAGAAATTAAGCCGAGACGTCGGTCGCGAATCTCGATTCGTATTA---AATATGCTTACT 5712  
QY 5676 CAATCAACTCTCCGAGCCAAATTTGTTTACTTATCTTCCGCTTACGAGATCCTGTGCAGC 5735  
Db 5713 CAATCACTTCTCCATCGCAATTTGTTGTTTGTGTCATCTGTATGGCTGACCTATAGAT 5772  
QY 5736 TGATCAATCTGTGTACAAATGCATTTGGTAAACCGATTTCAAACGCAACAGCTAGACAA 5795  
Db 5773 TGTAAACGTTTGTACAAATTCGTTAGGTAACCGATTTCAAACACAGCAAGCAAGACTA 5832  
QY 5796 CAGTCCAAACAGCAATTTGCGGATGCTCGAAACCTGTGCTAGTATGACAGTGAGATTTTC 5855  
Db 5833 CTGTTCAACAGAGTTTACGCGAGGTGTGGAACCTTTCCCTCAGAGCACCGTCAGATTTTC 5892  
QY 5856 CTGCATCGGATTTCTATGTGTATAGATATAAATTCGACGCTTGATCCGTTGATCACGGCT 5915  
Db 5893 CTGGCGATGTTTATAAGGTGTACAGGTACAAATGCAGTTTATAGTCTCTAATTTACTGCT 5952  
QY 5916 TATTAAATAGCTTTGATCTAGATAAATAAGATAATAGAGTTGATATCAACCCGACCGA 5975  
Db 5953 TGCTGGGGGCTTTTGATACTAGGAATAAGATAATCGAAGTAGAATAACCCAGCAGAGTCCGA 6012  
QY 5976 ATACTACTGAATCGTTTAAACGCGACTCAGAGGTTAGACGATGCTACTGTAGCTATAAGG 6035  
Db 6013 CAACAGCTGAACCGTTAGATGCTACCCGAGGTTAGACGACGCTACGTTGCAATTCGCT 6072  
QY 6036 CTTCAATCAATTAATTTGGCTAATGAACCTGTTCTGGAACTGGCATGTTCAATCAAGCAG 6095  
Db 6073 CTGCTATAAAATAATTTAGTTAATGAACCTAGTAAGAGGTTACTGGACTGTACAATCAGAATA 6132  
QY 6096 GCTTTGAGACTGTAGTGGACTTGTCTGGACCAACCTCCGCTACTTTAGCTATTTGTTGT 6155  
Db 6133 CTTTGAAGATGATGCTGGGTGTTGCTGGACCTCTGCACCTTGCAATCTTTAAATGCAATGCT 6192  
QY 6156 GAGATTTCTCAAAATAAAGTCGCTGAAGACTTAAATTCAGGGTGGCTGATACCAAAATC 6215  
Db 6193 GCTGAATATAAA-----GTTTGTGTTTCTAAACACACAGTGGTACGTAACG 6244  
QY 6216 AGCAGTGGTGTTCGTCACCTTAAATATAACGATTTGTCATATCTGGATCCCAAGTTAAA 6275  
Db 6245 TACAGTGTGTTTCTCCCTTAAATCGAAGGTTAGTGTCTTTGGAGCGCGGAGTAAAC 6304  
QY 6276 CCATGTGATGTTATCTACTGTGTTATGCGGTAAACATCGGAGAGGTTGCAATCTCTCCC 6335  
Db 6305 ATATATGGTTCATATATGTCGTTAGGACGCTAATAAAAAAGCGAGGATTTGCAATCTCCC 6364  
QY 6336 TAACCGCGGTAGCGGCCCA 6355  
Db 6365 GAAACCCCGTTGGGGCCCA 6384

RESULT 5  
AAZ20642  
ID AAZ20642 standard; RNA; 6395 BP.  
XX  
AC AAZ20642;















Db 2838 UUGCGGAAAUUGCAUUGUUUACGAGACACACACAGAGAUUCCAUCAUACAUAAGAUU 2897  
Qy 2885 CAGATTTCCTGATCCCAACATTTTCAGAGCTGCAAGTGAATGAGATTGAGATGAGG 2944  
Db 2898 UCAGAUUCCCGUACCCCGCCCAUUGGCAUUGGAAUUGGAGGAGGAGGACACGC 2957  
Qy 2945 AGGACACACATGAGATGCCAGGTGATGTGAATTTTCTCAATCGAAGTACGAAGGA 3004  
Db 2958 AGAACUACUCCUGUUGCCAGCGCAUGUCACACAUUACUGAACAGGAGAUUAGGGGC 3017  
Qy 3005 GGGGTGACAAACACATTCATCTGATCAACGATCGGTCTCATCTGAGATGATAGGCGGTAG 3064  
Db 3018 UUGUCAUGAGACACUUCUUGGUUAAAAGUGUUGUCGAGGAGUGGUGCGGCGAGCC 3077  
Qy 3065 GGGATCTAAACAGTGTTCACAAACACCTAAAGGGAATAATGTAACTTTCACTCAGGCT 3124  
Db 3078 GCGGUAUACAUCGACUACAACCCUUGCAUGGCAAGAUCCUGACUUAUCCCAUUG 3137  
Qy 3125 GATAAATTGAGTTAGAGAGAGGCTATAGAATGTGAACACCGTTTCATGAGATCCAA 3184  
Db 3138 GAUAAAGAGACUCUGCUUUAAGAGGUAUUCAGAUUGUACACUGUGCAUGAAGGCAA 3197  
Qy 3185 GGAGAAACCTTTGAAGATGTGCTGTGATGACGGCAACTCCACTGACTCTGATT 3244  
Db 3198 GCGGACAUACUCUGAUGUUUCAUAGUAGGUUAAACCCUACACAGUCUCCCAUCAU 3257  
Qy 3245 TCCAGTCTTCCCGCATGTTCTAGTCTGCTAGACACACAAAGAGCTTCAATAT 3304  
Db 3258 GAGAGACAGCCCAUGUUUUGUGCAUGUACAGGCCACACUGUUGCUCUAGUAC 3317  
Qy 3305 TACCCGTAGTTAGATCTTTAGTACAGATAATAGTAGATTGTCTTTTAAAGTCC 3364  
Db 3318 UACACUGUUGUAUGAUGUCCUUAUAGUUAUUAUAGAGUACUAGAGAAACUUAAGCUG 3377  
Qy 3365 TTCTCTTTAGAAATGTATGTAGTACAGAGTAGTAGATACGATATACAGATGATGCA 3424  
Db 3378 UACUUGUUAUUAUGUAAGUUGUUGAUGGAGGAAACAAUAGCAUUAUACAGAUUGCUG 3437  
Qy 3425 GTGTTCAAGGTCTAATCTCTTTGTGGCAACACCTTAATCAGGAGACTTCCAGATCTA 3484  
Db 3438 GUGUCAAAGGUUCCAAUUCUUUUGUGCAGCGCCCAAGACUGGUAUUAUUCUUAUG 3497  
Qy 3485 CAGTTCTATTAAGATGTATGCTCTCTGTTAATAGTATATATCTTAACAGATGATGCT 3544  
Db 3498 CAGUUUAUUAUGUAUAGUGUUGUCCAGGCAACAGCACCAUGAUGAAUUAUUAUGU 3557  
Qy 3545 GTTACCATGAGTTACGTGATATAGTCTTAATGTGAAGGATGTGTTCTGATTTTCC 3604  
Db 3558 GUUACCAUGGUGUACUGACAUUUCAUUGAUGUCAAAGAUUGCAUUAUGGAUUGUCU 3617  
Qy 3605 AAAAGTATTCGATGCCAAGGAGGTGAACCATGTCTAGAGCCAGTTTTCGTTACCGCG 3664  
Db 3618 AAGUCUUGUGCGCCUAGGAUCAAUUAACACACAUUACCUUAGUUGUACGAAACGGCG 3677  
Qy 3665 GCGGAACCGCGAAGGCTCGAGATCTCTGAAATCTGTTGCAATGATTAAGAAT 3724  
Db 3678 GCAGAAUUGCCAGCGCAGACUGGACUUAUUGGAAAUUUAUGGCGCAUGAUAUAAAGAA 3737  
Qy 3725 TTCAACGCCACGACCTGACGGGAGATTGTGATTTGAGAGACCGCATCTGTTGTAGTA 3784  
Db 3738 UUAACGACACCGAGUUGUGCAUCAUUGAUUUGAAUUAUUGCAUUCUUGGUUGUA 3797  
Qy 3785 GATAAGTTTTTATAGCTATTTTATTAAGAAAGAAAAATACACAAAAATATGCTGGA 3844  
Db 3798 GAUAAGUUUUUAUUAUUGUUAUUGUUAUUGAAAGAAACCAAAUAAAAUUGUUUCU 3857  
Qy 3845 GTGATGACAGAGATTCAATGATGATGATGTTGGAAAAACAGGAAGAAGTACTATTGGAC 3904  
Db 3858 UUGUACAGUAGAGUGUCUCUAUAGAUUGGUUAGAAACGAGAACAGGTAACAAUAGGC 3917  
Qy 3905 GACTTGGCTAACTACAAATTTTACAGATCTCGCGGCGCATCGATCAGTACAGACACATGATC 3964  
Db 3918 CAGCUCGCAUUAUUGAUUUUGGUAUUGCCAGCAGUUGAUCAGUACACAGACAUU 3977

Qy 3965 AAGGCTCAACCAACCAAGAAATTCGACCTTTCAATTAGAAATGAATACCTGCTCTGCAA 4024  
Db 3978 AAAGCAACAACCAACCAAGUUGGACACUUAUCCAAACGAGUACCCGCGCUUUGCAG 4037  
Qy 4025 ACAATTGCTACCAATTCGAAGCAGATCAACGGTATTTTGGCCGGTTTCTCA--GAGCTT 4081  
Db 4038 ACGAUUGUGUACCAUUCANAAAAAGAUCAUUGCAUUAUUCGGCCCGUUGUUGAGUGAGCU 4097  
Qy 4082 ACAAGGTTGCTGCTCGAGGCAATTTGATTTCTAAGAAATTTCTTTCTTTCTAGGAAACT 4141  
Db 4098 ACUAGCAUUAUUGGACAGUUGAUUUGGAGCAGAUUUUUUUUUCACAAGAAGACA 4157  
Qy 4142 CCAGAACAGATTTCAAGAAATTTTTCGCGATCTCGATCTCGACAGTTCCTATCGATGTGTTA 4201  
Db 4158 CCAGCGCAGAUUGAGGAUUCUUCGGAGAUUCUCGACAGUCAUGUGCGGAGUAGUCUUG 4217  
Qy 4202 GAACTGATATTTCTAAGTATGATAAGTTCACAGAACGATTTTCATTGTCGTAGAGTAT 4261  
Db 4218 GAGCUGGAUUAUCAAUAUACGACAAAUUCAGAAUUAUUCACUGUGCAGUAGAUAC 4277  
Qy 4262 GAATATGAAAAAGATTTGGGTCTCAATGAGTTTTCGCGGAGGTGTGAAAAACAAGGCAC 4321  
Db 4278 GAGUUCUGCGAGAUUGGUGUUAAGACUUCUUGGAGAGAUUUGGAAACAAGGCAU 4337  
Qy 4322 AGGAAAAACAATTGAAGGATTACATTGCTGGAATCAAGACATGCTGTGTATCAAAAG 4381  
Db 4338 AGAAAGACCAUCCAGGAUUAUACCGCAGGAUAAAAACUUGCAUCUGGUUAUCAAAGA 4397  
Qy 4382 AAAAGCGGTGATGATCTACTTTTCATCGCAATCTGTTAATAATAGACGTTGCTTGGT 4441  
Db 4398 AAGCUGGGAGCGUCACGACUUCUUGGAAACACUGUGAUCAUUGCUGCAUUGUUGGCC 4457  
Qy 4442 TCAATGTTTACCGATGAAAAAGGTCTATAAAGGTGCTTTTGTGGAGACGATTCGTTTTG 4501  
Db 4458 UCGAUCUCCGAGGAGAGAAUUAUCAAAGGAGCCUUUUGCGUGAGCAUAGUCUGCUG 4517  
Qy 4502 TATTTTCAAAAGGTTTGGATTTCCCTGACATTCAGTCAATGCTGCTAATCTCATGTGGAAT 4561  
Db 4518 UACUUCCAAAGGUGUGAGUUCGGAUGUGCAACUCGCGCAUUAUUGGUAU 4577  
Qy 4562 TTTGAGGCCAACTGTATAGAAAGAGGTACGGTACTTTTGTGGTAGATACATCATACAC 4621  
Db 4578 UUUUAGCAAAACUGUUUAAAAACAGUUGAUUAUUAUUGGAGAGAUUUGAUAUACU 4637  
Qy 4622 CATGATAAGGAGCAATAGTGTATATGATCTTTGAAAGTGTATCTCAAACTTGGGCA 4681  
Db 4638 CACGACAGAGAUUGUUGUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAU 4697  
Qy 4682 AAACATATCAAGGATTTATGATCATTAGAGTTAAGGTGCTTTTGTGCGATGTTGCT 4741  
Db 4698 AAACAUAUAGGAGUUGGAAACAUUGGAGGAGUUCAGAGGUCUUCUUGUAGUUGUCU 4757  
Qy 4742 TGTTCGCTCGGAAACTGCTGCTTAGCTTTTCGAGCTGAAACGACGATCTCAAGGAGGTT 4801  
Db 4758 GUUUCUU--GAACAUAUUGGUGUAUUAACACAGUUGGAGCGCUGUAGUGGAGUU 4814  
Qy 4802 CATAAAAACCGGATTTGATGTTGCTTTTAAATGTTGTTAAACAAATTTTGTGTGAT 4861  
Db 4815 CAUAAAGACCGCCUCCAGGUCUUGUUGUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAU 4874  
Qy 4862 AAATTTTATTTAGAACTTTGTTTAAATGCTGTAGTCTCAGAGATCTGTCAAAAT 4921  
Db 4875 AAAGUUCUUUUAAGAUUGUUUAUAGUGGUCUCUUGUUUUUAAGAAAGAAAGUAGAU 4934  
Qy 4922 TAGCGAGTTCAATGATCTTTTCAAAACAGGATGAGATATCTTCGCGCATTCATGACTAAGT 4981  
Db 4935 CAUAGAUUUAUCGACUUGCAAAUUAUGGAGAGAUUAUUAUUAUUAUUAUUAUUAUUAU 4994  
Qy 4982 CAAGAGTGTAGAAATATGATCTGTGGAACAAGATTATGGCTGTGTAAAGATGATGTTTC 5041  
Db 4995 AAAGAGUGUUAUGUGUUCCAAAGUUGAUAUUAUUAUUAUUAUUAUUAUUAUUAUUAU 5054





QY 1934 AAAAGCAGCGCTCTGAGCGCGTGTGTATGTCTTGAAACCGACATCCGAAGAGGTGAACGTA 1993  
Db 1938 CAAGAGAAGGCTTCAGAAAGGTGCTTTGGTGTAGTTACCTCAAGAGAAGTTGAAGAACCGGTCC 1997  
QY 1994 AATAAAATTTCTATTGCTGAGAAAGGAGATTGCCGTGTGTGCGAAGATCATGGTTTG 2053  
Db 1998 ATGAAGGGTTCATGCGCCAGAGGAGATTCAATTAGCTGGTCTTTGCTGGAGATCATCCG 2057  
QY 2054 ACGAATGCTAACTTAGAGCACAGAGATTGGAGTCCCTCAACGATTTCCATAAGGCTTGC 2113  
Db 2058 GAGTCGTCTTATTCTAAGAAACGAGAGATAGAGTCTTTAGAGCAGTTTCATATGCGAAG 2117  
QY 2114 GTGGATGTGTATTACAAAGCAATGGCATCGGTTGTCTACATGCGCTCACTCAAGTTT 2173  
Db 2118 GCAGATTCTGTTAAATTCGTAAGCAGATGAGCTCGAATTGTGTACACGGGTCCGATTAAGTT 2177  
QY 2174 CAACAAATGAAGAATATGTGACAGTTTGGCAGCTTCTGTCGCGCACCTGTATCAAAAT 2233  
Db 2178 CAGCAATGAANAACCTTTATCGATAGCTGTGTAGCATCACTATCTGCTGCGGTGTGCAAT 2237  
QY 2234 CTATGCAAGTCACTAAAGGATGAAGTCCGGGTATGATTTCTGATTTCCAGGGAGAAAGTTGGT 2293  
Db 2238 CTGCTCAAGATCCTCAAGATACAGCTGCTATTGACCTTGAACCCCGTCAAAAGTTTGA 2297  
QY 2294 GTTTGGGATGTCACTTTGAAAGGTGGCTCCTCAAAACCTGCGGCCCAAGGTCATTCATGG 2353  
Db 2298 GTCTTGGATGTGTCACTCTAGGAAGTGTGTTAATCAAAACCAACGGCCCAAGAGTCATGCTGG 2357  
QY 2354 GGAGTTGCTCGGATTACAAAGGGGAAATGTTTACTGCACTTCTATCTTATGAAGGAGAT 2413  
Db 2358 GGTGTGTGTGAACCCACCGAGGAAAGTATCATGTGGCGCTTTTGGAAATATGATGAGCAG 2417  
QY 2414 AGAATGTGTACTGAGAGCGACTGGAGGAGGGTGTGTATCATCTGATACAAATGGTATAT 2473  
Db 2418 GGTGTGTGACATGCGATGATTGGAGAGAGTAGCTGTGAGCTCTGAGTCTGTTGTTTAT 2477  
QY 2474 TCTGATATTGCAAGCTCAAAATCTGAGGAAACCAATGAGAGCGGTGAACCCCAAGAA 2533  
Db 2478 TCCGACATGGCGAAACTCAGAACTCTGCGCAGACTGCTCGAATTCGAAACCGAGAACCCGATGTC 2537  
QY 2534 CCTCTGCAAGATGTTGATGAGTGGGTGCTGCTGTGTGGAACTACAAAGGAGAT 2593  
Db 2538 AGTAGCGAAAGTTGTTCTGTGAGCAGAGTTCGGGCTGTGGGAAACCAAGAAAT 2597  
QY 2594 TTTGAAAGATTGATCTTGATGAGGATTGATCTTGTTGTTCTCGGAAACAAAGCTGCTGCT 2653  
Db 2598 CTTTCCAGGGTTAAATTTGATGAAGATCTAAATTTAGTACCTTGGGAACCAAGCGGAA 2657  
QY 2654 ATGATCAGAAGAGGGCTTAATCTATCTGGAATGATGATGAGGCAATGGAACAATGTGAGA 2713  
Db 2658 ATGATCAGAAGACGTGCGAATTTCTCAGGGATTTATGTGGCCACCAAGAGCAACGTTAAA 2717  
QY 2714 ACGGTAGATTCACTTCTAATGAT-----CCAAACCCGATCACAGAGGCTT 2764  
Db 2718 ACCGTTGATCTTTTCAATGATGAATTTGGGAAACGACACGCTGTGAGTTCAAGAGGTTA 2777  
QY 2765 TTTATTGATGAAGGGTTGATGCTGACACCGGTTGTGTAACTTCTCGGTGCTTATCTCT 2824  
Db 2778 TTCAATTGATGAAGGGTTGATGCTGATCTACTGTTGTGTTAAATTTTCTTGTGGCGATGTA 2837  
QY 2825 GGTTCGCAATCGCATACATTTACGAGATACACAGCAGATTCCTTTTCAATTAACAGATT 2884  
Db 2838 TTGTGCGAAATGCAATATGTTTACGAGACACACAGCAGATTTCCATACATCAATAGAGTT 2897  
QY 2885 CAGAATTTCCGATATCCCAACATTTTCGAGAGCTGCAAGTGGATGAAGTGGATGAGG 2944  
Db 2898 TCAGGATTTCCCGTATCCCGCCCATTTTTCGCAATTTGGAAGTTGAGAGTTGGAGACACGC 2957  
QY 2945 AGGACACACTGAGATGCGCAGGTGATGTGAATTTTTTCTTACAACTCGAAGTACGAAGGA 3004  
Db 2958 AGNACTACTCTCCGTTGTCCAGCCGATGTCACACATATCTGACAGGAGATATGAGGGC 3017

QY 3005 CGGTGACCAACCACTTCAACTGTACACGATCGGTCTCATCTGAGATGATAGGCGGTAAAG 3064  
Db 3018 TTTGTCATGACACTTCTTCGTTTAAAAAGTCTGTTCGCGAGAGATGGTCGGCGGAGCC 3077  
QY 3065 GGAGTACTAAACAGTGTTCACAAACCACTAAAGGGAATTTGTAACCTTTCACTCAGGCT 3124  
Db 3078 GCGGTGATCAATCGATCTCAAAACCTTGCATGGCAGATCCTGACTTTTACCCAAATCG 3137  
QY 3125 GATAAATTTGAGTTAGAGGAGAAGGCTTATAGAAATGTGAACACCGTTTCATGAGATCCAA 3184  
Db 3138 GATAAAGAAGCTCTGCTTTCAAGAGGGTATTTAGATGTTTCACTGTGCTGAAGTGCAG 3197  
QY 3185 GGAGAAACCTTTGAAGATGTGCTGCTGATGAGGCAACCTCCACTGACTCTGATTT 3244  
Db 3198 GCGGAGACATCTCTGATGTTTCACTAGTTAGGTTAACCCCTACACCCAGTCTCCATCAT 3257  
QY 3245 TCCAAGTCTTCCCGCATGTTCTAGTCTGCTCTGACTAGACACACAAAGAGCTTTCAAAAT 3304  
Db 3258 GCAGGAGACGCCACATGTTTGGTCGATTTGTCAGGACACCTGTTGCTCAAGTAC 3317  
QY 3305 TACACCGTGTGTTAGATCTCTTTTAGTACAGATAATTTAGTGTGTTCTTTTAAAGTCC 3364  
Db 3318 TACACTGTTGTTATGAGTCTCTTTAGTTAGTATCATTTAGAGATCTAGAGAACTTAGCTCG 3377  
QY 3365 TTCTTTTGTAGAAATGATATGTTAGGACGAGTAGATAGCAATTTACAGATGGATGCA 3424  
Db 3378 TACTGTTGTAGATATGTTAAGTGTGATGAGGAAACAAATAGCAATTTACAGATGACTCG 3437  
QY 3425 GTGTTCAAAGTGTATATCTCTTTGTGGCAACACCTTAATCAGGAGACTTTCCAGATCTA 3484  
Db 3438 GTGTTCAAAGTGTCCATCTTTTGTGCGGCCAAAGACTGTGTGATATTTCTGATATG 3497  
QY 3485 CAGTCTTATACGATGTATGCTCCCTGGTAAATAGTACTATATCTTAAACAAGTATGATGT 3544  
Db 3498 CAGTCTTACTATGATAGTGTCTCCAGGCAACAGCACCATGATGAATAAATTTTGTATGCT 3557  
QY 3545 GTTACCAGTGTACGTGATAATAGTCTTAATGTGAGGATGTGTTCTTGTGATTTTCC 3604  
Db 3558 GTTACCAGTGTGTGACTGACATTTTCAATTGAATGTCAAAGATTCATATTTGATATGCT 3617  
QY 3605 AAAAGTATTCGATGCCAAGGAGGTGAACCATGTCTAGAGCCAGTTTTCGTACCGCG 3664  
Db 3618 AAGTCTGTTGTCGCTTAGGATCAATCAACCACTAATACCTATGTTAGTACGACGCG 3677  
QY 3665 GCGAAACCGCAAGGGTGTGAGGACTACTCGAAATCTGTTGCAATGATTAAGAAAGAT 3724  
Db 3678 GCAGAAATGCCACGCGAGACTGGAATAATGGAATAATTTAGTGGCGATGATTAAGAAAGAAC 3737  
QY 3725 TTCAACGACACGACCTGACCGGGACGATTCACATTTGAGAGCACCGCATCTGTTGATGA 3784  
Db 3738 TTTAACGACCCGAGTTGTCTGGCATCATTTGATATTTGAATAATCTGTCATCTTTAGTTGA 3797  
QY 3785 GATAAGTTTTTTGATGATCTATTTTATAAAGAAAGAAATACAAAAAATATTCGTGGA 3844  
Db 3798 GATAAGTTTTTTGATGATTTTGTCTTAAGAAAGAAAGAAACCAATATAAATGTTTCT 3857  
QY 3845 GTGATGACGAAGGATTCATGATGATGATGTTGGAAGAAACAGAAAGAGTACTATTTGAGAC 3904  
Db 3858 TTGTTTCAGTAGAGTCTCTCAATAGATGTTTAGAAAGAGGAGAACAGTAACAATAGGC 3917  
QY 3905 GACTTGGCTACTACAAATTTTACAGATCTGCGGCCATCGATCGATCAAGACACATGATC 3964  
Db 3918 CAGCTCGCAGATTTTGTATTTTTTGTAGTTTTGCGAGCTGTGATTCAGTACAGACATGAT 3977  
QY 3965 AAGCTCAACCAAAACAGAAATTCGACCTTTCAATTCAAGATGAATACCTGCTCTGCA 4024  
Db 3978 AAAGCAAAACCAAGCAAAATTTGACACTTCAATCCAAACGAGTACCCGGCTTTGAC 4037  
QY 4025 ACAATTTGCTACCAATCGAAGCAGATCAACGGTATTTTGGCGGTTTCTCA---GAGCTT 4081  
Db 4038 ACGATTGTGTACCAATCAAAAAAGATCAATGCAATATTTGGCCCGTTGTTAGTGAGCTT 4097  
QY 4082 ACAAGTTGTGCTCGAGGCAATTTGATTTCTAAGAGATTTCTTTTCTTACTAGGAAACT 4141

Db 4098 ACTAGGCAATTAAGTGGACAGTGTGATTCGAGCAGATTTTGTGTTTTCACAAAGACACA 4157  
Qy 4142 CCAGAACAGATTCAGAAATTTTCTCGGATCTCGACTCGCAGTTTCTTATGATGATGTTA 4201  
Db 4158 CCAGCCAGATTTAGAGATTTCTTCGGAGATCTCGACAGTCATGTGCCGATGATGCTTG 4217  
Qy 4202 GAACCTGGATATTTCTAAGTATGATAAGTCACAGAACGAGTTTCAATGTGTGCTGTAGAGTAT 4261  
Db 4218 GAGCTGGATATATCAAAATACAGCAAACTCAGATGAATTTCCACTGTGCAGTAGAATAC 4277  
Qy 4262 GAAATATGGAAGAAAGTGGTCTCAATGAGTTTTTGGCCGAGTGTGGAACCAAGGCAAC 4321  
Db 4278 GAGATCTGGCGAAGATTTGGTTTGAAGACTTCTTGGGAGAGTTTGGAAACCAAGGCAT 4337  
Qy 4322 AGGAAACCAACTTTGAGGATACATTTCTGGAATCAAGACATGCTGTGGTATCAAGG 4381  
Db 4338 AGAAAGACACCCCTCAAGGATATACCCGAGGTATAAAACCTTGCATCTGGTATCAAGA 4397  
Qy 4382 AAAAGCGGTGATGACTACTTTTCATCGCAATACCTGTTATATAGCAGCTTGTGCGT 4441  
Db 4398 AAGAGCGGGACGTCACGAGTTTCATTTGGAACACTGTGATCATTTGCTGCAATTTGGCC 4457  
Qy 4442 TCAATGTTTACGATGGAAGAGTCTAATAAGGTGCTTTTGTGAGAGACGATTCGCTTTTG 4501  
Db 4458 TCGATGCTTCGATGGAGAAATAATCAAGGAGCCTTTTGGGTGACGATAGTCTGCTG 4517  
Qy 4502 TATTTTCCAAAGGGTTTGGATTTCCCTGACATTTCAATGCTAGTCTTAATCTCATGTTGAAT 4561  
Db 4518 TACTTTCCAAAGGGTTGAGTTTCCGATGTGCAACCTCCGGAATCTTATGTGGAAT 4577  
Qy 4562 TTTGAGGCCAAACTGTATAGAAAGAGTACGGTTACTTTTGTGAGTAGATACATCATACAC 4621  
Db 4578 TTTGAGCAAACTGTTTTAAAAAAGATGATGATGATTTTGGGAGATATGATATACAT 4637  
Qy 4622 CATGATTAAGGAGCAATAGTGTATATGATTCCTTTGAGTTGATCTCMAACTTGGGCA 4681  
Db 4638 CACGACAGAGGATGATTTGTGATTTACGATCCCTAAAGTTGATCTCGAAACTTGGTGT 4697  
Qy 4682 AATCATATCAAGGATTTATGATCACTTTAGAAAGATTTAAGGTTCTTTTGGCGATGTTGT 4741  
Db 4698 AATCATATCAAGGATTTGGAACACTTTGAGGAGTTTCAAGATCTCTTTGTGATGTTGT 4757  
Qy 4742 TGTTCGCTCGAAACTGTGTGCTTAGGCTTTCCGACGCTGAAACGACGATATCAAGAGTT 4801  
Db 4758 GTTTCGTT--GAACAAATTGTCGATTTACACACAGTTTGGACGACGCTGATGGAGTT 4814  
Qy 4802 CATAAACCGGATTTGATGTTGCTTTGCTTTTAAATTTGTTTAAACAAATTTTGTGTTGT 4861  
Db 4815 CATAAAGACCGCCCTCCAGGTTTCTGTTTGTATTAAGTCTGGTGAAGTATTTGTCTGAT 4874  
Qy 4862 AAATTTTATTTAGAACTTTGTTTAAATGCTGTAGTCTCAGAGATCTGTCAAAAT 4921  
Db 4875 AAGTTCTTTTGAAGTTTGTATATGATGCTCTAGTTGTTAAAGGAAAGTGAATAT 4934  
Qy 4922 TAGCGATTTCAATGATCTTTTGAAGACAGGATGAGATCTTCCGGCATTCATCACTAAGT 4981  
Db 4935 CAATGAGTTTATCGACTTCACAAAATGGAAGATCTTACCGTTCGATGTTTACCCCTGT 4994  
Qy 4982 CAAGAGTTTAGAATATCACTGTGGACAAAGTATATGCTGTGTTAAGATGATGCTTTC 5041  
Db 4995 AAGAGTGTATGTGTTCCAAAGTTGATATAAATAATGTTTCATGAGAATGATGCTATGTC 5054  
Qy 5042 TGATGTAGATTTTACTTTAAAGGTGTTAAGTTAGTTAAGAAAGGATGATGCTGTAGCTGA 5101  
Db 5055 AAGAGTGAACCTTCTTAAAGGAGTTAAGCTTATGATAGTGAATGATGCTGTGTAGCCGG 5114  
Qy 5102 TTTGTGATGTGTGGGAGTGAATCTCCGGATATCTCCGCTGGTGTGCTGAGTGTGTTG 5161  
Db 5115 TTTGTGCTGTCACGGGCGAGTGAACCTTGCCTGACATTTGAGAGGAGTGTGAGCGTGTG 5174  
Qy 5162 TATTTAGATTAAGAGATGAAGAGGATGAAGNAGCAACGCTGGGTGCTATCAGCCGCC 5221

Db 5175 TCTGGTGACAAAAGGATGGAAGAGCGGACGAGGCCACTCTCGGATCTTTACTACACAGC 5234  
Qy 5222 TGCTTGCAAAAAGAAATTTTCTTTAAAGCTAATCCCTAAATTAATCAATAAACATCCGAGGA 5281  
Db 5235 AGCTGCAAGAAAGATTTTCAAGTTCAAGGTCGTTCCCAATTAATGCTATATACCAACCCAGGA 5294  
Qy 5282 TGCTGAGAAAGCAACCGTGGCAAGTGTGTAGTGAATATCAAGAGGATGCTGTGGAAGAGG 5341  
Db 5295 CGCGATGAAAACCGTGGCAAGTGTGTAGTGAATATTAAGAAATGTGAAGATGTGAGCGGG 5354  
Qy 5342 ATACTGCTCTTATCTTTTGGAGTTTCGTTCAATTTGTGTAGTACATAAAATAATATGTAAG 5401  
Db 5355 TTTCTGCTCCGCTTTCTTGGAGTTTGTGTCGGTGTGTATTTTATAGAAATAATATATAA 5414  
Qy 5402 AAAAGGTTTGAAGGAAACGTAATTTTGAAGTGTGACAGACGCGCTCGCCAAATTTGAACCTCAGTGA 5461  
Db 5415 ATTAGTTTGAAGAGAGATTTACAAAGCTGAGAGACGAGGCGCCATGGAACCTTACAGA 5474  
Qy 5462 AAAAGTTTGAAGAGTTTGTGATGAAGTACCAATGCTGTGAAAATCTGGAAGAGTTTC- 5520  
Db 5475 AGAAGTCTTGTGATGAGTTTCAAGAGATGTCCTATGTGTCGATCAGGCTTGCAAGTGTTCG 5534  
Qy 5521 -----CGGAACAAACAAAGAAATGCTAGGTAATAATGTTAATAAT----- 5561  
Db 5535 ATCTGAAACCCGAAACAAAGAGTGTCCGCAAGGGAATAATAGTAGTAATGATCGGTC 5594  
Qy 5562 -----AAGAAAATAATAACAGTGTGTAAGAGGGTTTTTAAATTTGAGGAAATTTAGGA 5614  
Db 5595 AGTCCGCAACAAAGTATAGAAATGTTAAGATTTTGGAGGAATGAGTTTTTAAAGAA 5654  
Qy 5615 TAAATGTAAGTGAAGCAGAGT-----CTATCGCTCATCGAGTACGTTTAAATCAAT 5665  
Db 5655 TAAATTAATCGATGATGATTCGAGGCTACTGTGCGCGAATCGGATTCGTTTAA- 5711  
Qy 5666 ATGCTTATACATCAACTCTCCGAGCCAAATTTGTTTACTTATCTTCGCTTTACGACAT 5725  
Db 5712 ATGCTTACAGTATCACTACTACCTCCATCTCAGTTCTGTTCTTGTCTATCAGCGTGGGCCAC 5771  
Qy 5726 CCGTGCAGCTGATCAATCTGTGTACAAATGCAATGGGTAAACGATTTTCAACGCAACAA 5785  
Db 5772 CCAATAGAGTTTAAATTTAATTTATGATTAATGCTTTAGGAATCAGTTTCAACACAAACAA 5831  
Qy 5786 GCTAGGACAAACAGTCCAAACAGCAATTTCCGAGTCCGTAACCTGTCGCTAGTATGACA 5845  
Db 5832 GCTCGAATCTGCTTCAAGAACNAATTCAGTGAGGTGTGGAACCTTCCACCACAGTAAT 5891  
Qy 5846 GTGAGATTTCTGTCATCGATTTCTATGCTATAGATATAATTCGACGCTTGCATCCGTTG 5905  
Db 5892 GTTAGTTTCCCTGACAGTGAATTTAAGGTGTACAGGTACATTCGCTATTAGACCCGCTA 5951  
Qy 5906 ATCAGCGGTTTATTAATAGCTTTGATATCTAGAAATAGAAATAATAGAGGTTGATATCA 5965  
Db 5952 GTCCAGACACTGTTAGGTGCAATTCGACACTAGAAATAGAAATAATAGAAAGTTGAAATCAG 6011  
Qy 5966 CCCGACACGATATCTACTGAAATCGTTAAACGCTCAGAGGTTAGAGTCTACTGTA 6025  
Db 6012 GCGAACCCACGACTGCGCAACCGTTAGATGCTACTCTGTAGGTAGACGACCAACGCTG 6071  
Qy 6026 GCTATAAGGCTTCAATCAATAATTTGGCTAATGAACTGTTTCTGTTGGAACCTGGCTGTT 6085  
Db 6072 GCCATAGAGCGCGATTAATAATTTAATAGTAGAATTTGATCAGAGGAACCGGATCTTAT 6131  
Qy 6086 AATCAAGCAGGCTTTGAGACTGCTAGTGGACTTTGTCTGGAACCAACCTCCGCTACTTAG 6145  
Db 6132 AATCGAGCTCTTTCGAGAGCTCTTCTGTTTGGTTTGGACCTCTGCTGCTGCTCAACT--- 6188  
Qy 6146 CTATTTGTTGTGAGATTTCTTAAATAAAGTCTGCTGAAGCTTTAAATTTCAAGGTTGCTGA 6205  
Db 6189 -TGAGGTAGTCAAGATGATATAATAATACCGATTTGTGTCGTTATACCA-CGTGGTGGC 6246  
Qy 6206 TACCAAAATCAGCAGTGGTTGTTTCCTCCACTTTAAATATAACGATTTGTCTATCTCGATCC 6265  
Db 6247 TACGATACGCATAGTGTGTTTTTCCCTCCACTTTAAATCGAAGGTTGT-GTCTTGGATCGC 6305







1100 AAAACCTTGGCCATGTTCAACACTGAAAGAGCAATCTTTAGAGACACGGCTTCGGTTAAC 1159  
1098 AAGACTCTTGCAATGTGCAACACGCGAGAAATCCCTCTTGAGGATTCATCATCAGTCAAT 1157  
1160 TTTTGGTTCCTTAAGATGAAGACATGGTATGATGACCGCTGTTTGAGGGTTCATATTACC 1219  
1158 TACTGGTTTCCCAAATGAGGATATGATCATCGTACCATTTATTCGACATTTCTTTGGAG 1217  
1220 AGCAAAAAGATGACAAAGGATGAGTCAATGTTAATCGTGACTTCGGTTTACACAGTGTCT 1279  
1218 ACTAGTAGAGGACGCGAAGAAATCTTAGTGTCCAAGGATTTCTGTTTACAGTGTCT 1277  
1280 AATCATATCAGAACATATCAAGCCAAAGGTTAACTTACCAGAAAGTATTATCTTTCTGTG 1339  
1278 AACCACATTCGAACATACACAGCGAAAGCTTTATACATCGCAAAATGTTTGTCTTTGTC 1337  
1340 GAGTCTATAAGATCCCGGTGATAATCAATGCTGTTACTGCTAGGCTCTGAATGGGATGTA 1399  
1338 GAATCGATTCGATCGAGGTTATCATTAACGCTGTGACAGCGAGTCCGAATGGGATGTG 1397  
1400 GATAAGCAATTTCAACCCCTTGTCAATGACTTTCTTCTTCAGACTAAGCTGGCTGG 1459  
1398 GACAAATCTTGTCTCAATCTTGTCCATGACGTTTACCTGCATCTAAGCTTCCCGTT 1457  
1460 CTTCAAGACGATATAGTAATGGAAGTTTCGGTCTTGATAAGAACCACTTCTCAACTT 1519  
1458 CTAAGGATGACTTACTGATTAGCAATGTTTAGTCTCGGTTCCAAACCGTGTGCCAGCAT 1517  
1520 ATTTGGGATGAGTGGGCAAAATTTTGGAAACGTTTTTCCCACTATCAAGAGAGATG 1579  
1518 GTGTGGATGAGATTTCCGTGGCTTTGGGAACGCAATTTCCCTCGTGAAGAGAGGCTC 1577  
1580 GTGACGAGAAATTTCTGATGTAGTGAGATGCTCTCAAGCTCAAGATCCAGATCTG 1639  
1578 TTGAACAGGAACTTATCAGAGTGGCAGCGCAATTAGAGATCAGGGTGCCTGATCTA 1637  
1640 TATGTCACATGGAAGACAGGTTCTGATCTGTAATACCAAGTCTGAGGAGTTTACCGCAT 1699  
1638 TATGTGACCTTCCACACAGATTAGTGTACTGATGACAGGCTCTGTGACATGCTCGG 1697  
1700 CTAGATATCAAGAGGACTTAGAAGAGCTGAGCAAAATGTACGACCGGTTATCAGAAATTA 1759  
1698 CTTGACATTAGGAAGAGATGGAAGAAACGGAAGTATGTACAAATGCACTTTTACAGAGTTA 1757  
1760 TCTATCCTTAAGGGTCTGATATTTTCTGATATCGGAAGTTTCAAGACATGTGCAAGGCT 1819  
1758 TCGGTGTTAGGGAGTCTGACAAATTCGATGTTGATGTTTTCACGATGTGCCAATCT 1817  
1820 TTAGATGTTAGTCTCTGATGTGGCAGCACGAGTAAATCGTTGCAATGCGCGAGAAATAGAAGC 1879  
1818 TTGGAAGTTGACCAATGACGCGCAGCGAAGTTTATAGTCGCGTCTATGAGCAATGAGAGC 1877  
1880 GGTTTAACTTACTTTTATGATAAGCCACCGAGAGAAATGTGCTAAGGCTCT- ----T 1933  
1878 GGTCTGACTCTCAATTTGAACGACCTACTGAGGCGAATGTTGCGGTAGCTTTTACAGGAT 1937  
1934 AAAAGCAGCGCTCTGAGGCGGTGATGTTCTTGAACCGACATCCGAAGAGGTGAACGTA 1993  
1938 CAGAGAGGCTTTCAGAGGTGCTTTGGTAGTTTACCTCAAGAGAGTTTGAAGAACGCTCC 1997  
1994 AATAAATTTTCTATGCTGAGAAGGAGATTCGCTGTGTGTGTCAGAAAGTCAATGTTTG 2053  
1998 ATGAAGGGTTTCGATGGCCAGAGGAGATTAACAATTAGTCTGTTGCTGGAGATCATCCG 2057  
2054 ACNAATGCTAACTTAGACACAGGAGTTCGGAGTCCCTCAACGATTTCCATTAAGGCTTGC 2113  
2058 GAGTCGCTCTATTCTTAAGAACGAGGAGATAGAGTCTTTAGAGCAGTTTCAATATGGCAAG 2117  
2114 GTGATAGTGTGATTTACAAAGCAAAATGGCATCGGTTGTCTACACTGCTCACTCAAGTT 2173  
2118 GCAGATTCGTTAATTCGTAAAGCAGATGAGCTCGATTTGTGTACACGCGTCCGATTAAGTT 2177

2174 CAACAAATGAAGAACTATGTGGACAGTTTGGCAGCTTCGTTGTCGCCCACTGTATCAAAAT 2233  
2178 CAGCAATGAAGAACTTTATCGATAGCTTGGTAGCACTATCTCTGCTGGGTGTCGAAT 2237  
2234 CTATGCAAGTCACTAAAGGATGAAGTCGGGTATGATTTCTGATTTCCAGGGAGAAAGTTGGT 2293  
2238 CTCGTCAAGATTCCTCAAGATACAGCTGCTATTGACCTTGAAACCCGCTCAAAAGTTTGGGA 2297  
2294 GTTTGGGATGTCACTTTGAAAAAGTGGCTCTCTCAAACTCGCGCAAAAGGTCATTCATG 2353  
2298 GTCTTGGATGTCATCTAGGAAGTGGTTAATCAAAACCAACGCGCAAGAGTCATGCAATGG 2357  
2354 GGAGTTGTCTCGGATTAACAAGGGGAAATGTTTACTGCACCTTCTATCTTATGAAGAGAGAT 2413  
2358 GGTGTTGTTGAAACCCACGCGAGGAAGTATCATGTGGCGCTTTTGGAAATATGATGAGCAG 2417  
2414 AGAATGTTGATGAGAGGACTCGAGAGGGTGGCTGTATCATCTGATACATAATGTTATAT 2473  
2418 GGTGTGGTGAATCGGATGATTTGAGAGAGTAGTCTGTCTAGCTCTGAGTCTGTTGTTTAT 2477  
2474 TCTGATATTGCAAAAGCTCCAAAATCTGAGGAAAAACAATGAGAGACGGTGAACCCCAAGAA 2533  
2478 TCCGACATGGCGAAACTCAGAACTCTGCGCAGACTGCTTCGAAACGGGAGAACCGCATGTC 2537  
2534 CCTACTGCAAAAGATGTTACTTGTGGATGGGTGCTGTTGTGGAAGATCAAAAGGAGAT 2593  
2538 AGTAGCGCAAGGTTGTTCTTGTGGACGGAGTTCGGGCTGTGGGAAACCAAAAGAAAT 2597  
2594 TTTGAAAGATTTGATCTTGATGAGGATTTGATCTTGGTTCTCGGAAAAACAAGCTGCTGCT 2653  
2598 CTTTCCAGGTTAATTTTGTGTAAGATCTAATTTTAGTACCTTGGGAGCAAGCCGCGAA 2657  
2654 ATGATCAGAAGAGGGCTTAATCTGACTGATTAAGAGCCACAATGGAACAATGTGAGA 2713  
2658 ATGATCAGAAGACGTGCGAATTCCTCAGGGATTAATGTGGCCACGAGAGCAACGTTAAA 2717  
2714 ACGGTAGATTCACTTCTAATGCAT-----CCAAAACCGCATCACAAAGAGGCTT 2764  
2718 ACCGTGATTTTTCATGATGAATTTTGGGAAAAAGCACACGCTGTGAGTTCAAGAGTTA 2777  
2765 TTTATTTGATGAAGGGTTGATGCTGCACACCGGTTGTGTTAACTTCCTGGTCTCTATCTCT 2824  
2778 TTTATTTGATGAAGGGTTGATGTTGCATACTGTTGTGTTAATTTCTTGTGGCGATGCA 2837  
2825 GGTTCGCAATCGCAATCAATTTACGAGATACACAGAGATTCCTTTTCAATTAACAGAGTT 2884  
2838 TTGTGCGAAATTCATATGTTTACGAGACACACAGCAGATTTCCATACATCAATAGAGTT 2897  
2885 CAGAAATTTCCGTTATCCCAACATTTTCGAAAGCTGCAAGTGGATGAAGTTGAGATGAGG 2944  
2898 TCAGGATTTCCCGTACCCGCCCATTTTCCCAATTTGGAAGTTGACGAGTGGAGACACGC 2957  
2945 AGGACACACATGAGATGCCAGGTTGATGAATTTTTCCTACAATCGAAGTACGAAGA 3004  
2958 AGAATCTACTCTCGTTGTCACCGCATGTCTCACATTTATCTGAACAGAGATATGAGGCG 3017  
3005 GGGGTGACAAACCACTTTCAACTGTACACGATCGGTTCTCATCTGAGATGATAGCGGTAAG 3064  
3018 TTTGTGATGAGCACTTCTTCGGTTTAAAGTCTGTTTCGAGGAGATGTTGCGCGAGGCC 3077  
3065 GGAGTACTAAACAGTGTTCCTCAAAACCACTTAAAGGGGAAATTTGTAATTTTCACTCAGGCT 3124  
3078 GCGGTGATCAATCCGATCTCAAAACCTTTGATGCGCAAGATCTCTGACTTTTTACCAATCG 3137  
3125 GATAAATTTGAGTTAGAGAGAGGGCTTATAAGATGTGAACACCGTTCATGATGATCCAA 3184  
3138 GATAAAGAGCTCTGCTTTCAAGAGGGTATTTCAGATGTTTCACTGTGATGATGAAGTCAA 3197  
3185 GGAGAAACCTTTGAAGATGTCGCTGTCAGATTGACGGCAACTCCACTGACTCTGTGATT 3244  
3198 GCGGAGACATCTCTGATGTTTCACTAGTTAGTTAAACCCCTACACCACTCTCCATCAT 3257  
3245 TCCAAAGTCTTCCCGCATGTTCTTAGTCTGCTGTGACTAGACACAAAGAGGCTTCAAAATAT 3304

Db 3258 GCAGAGACAGCCCATGTTTGGTCGCAATGTCAGGCACACCTGTTGCGTCAAGTAC 3317  
Qy 3305 TACACCGTAGTGTAGATCCTTTAGTACAGATAAATAGTAGATTTGCTCTTTTAAAGCTCC 3364  
Db 3318 TACACTGTTGTTAGTATGATCCCTTTAGTGTAGTATCAATTAGAGATCTAGAGAACTTAGCTCG 3377  
Qy 3365 TTCCCTTTAGAAATGATATGTTAGTAGACAGGAGTAGATAGCAATTAACAGATGATGCA 3424  
Db 3378 TACTTGTGTAGATATGATATAGGTCGATGACGAAACAATAGCAATTAACAGATGACTCG 3437  
Qy 3425 GTGTTCAAGGTCATAATCTCTTTGTCGCAACACCTAAATCAGGAGACTTTCCAGATCTA 3484  
Db 3438 GTGTTCAAGGTTCCAAATCTTTTGTGACGCCAAAGACTGGTGATATTTCTGATAG 3497  
Qy 3485 CAGTTCTATTACGATATGCTCCCTGGTAAATAGTACTATATCTTAAACAAGTATGATGCT 3544  
Db 3498 CAGTTTACTATGATAAGTGTCTCCAGGCAACAGCACCATGATGAATAATTTTGATGCT 3557  
Qy 3545 GTTACCATGAGGTTACGTGATATAATAGTCTTAATGTGAAGAGATGTGTTCTTGTATTTTCC 3604  
Db 3558 GTTACCATGAGGTTGACTGACATTTCAATTGAATGTCAAAGATTGCATATTGGATATGCT 3617  
Qy 3605 AAAAGTATTCGATGCCAAGAGGTGAACCATGTCTAGAGCCAGTTTTCGTACCGG 3664  
Db 3618 AAGTCTGTGTGCGCCCTAAGGATCAAAATCAACCACTAATACCTATGGTACGAACGGCG 3677  
Qy 3665 GCGGAACCGCCAAAGGCGTCGAGGACTACTCGAAAAATCTGGTTGCAATGATTAAGAAAT 3724  
Db 3678 GCAGAAATGCCACGCCAGACTGGACTATTTGGBAAATTTAGTGGCGATGATTAAGGAAAC 3737  
Qy 3725 TTCAACGCACAGACTCGACGCGGAGCATGTGACATGTGAGAGACACCGCATCTGTTGTAGTA 3784  
Db 3738 TTTAAGCAGCCCGAGTTGCTGGCATCATTTGATATTGAATAATCTGTCATCTTTAGTTGTA 3797  
Qy 3785 GATAAGTTTTTGTAGTCTATTTTATTAAGAAAGAAATACACAAATAATTTGCTGGA 3844  
Db 3798 GATAAGTTTTTGTAGTATTTGCTTAAAGAAAGAAACCAATTAAGAAATGTTTCT 3857  
Qy 3845 GTGATGACGAAGGATTCATGATGATGATGGTTGGAAAAACAGAAAGAGTACTATTGGAC 3904  
Db 3858 TTGTTCAGTAGAGTCTCTCAATAGATGTTAGAAAGCAGGAACAGTAAACATAGGC 3917  
Qy 3905 GACTGGCTAACTACAATTTTACAGATCTGCGCGCATCGATCAGTACAGACACATGATC 3964  
Db 3918 CAGCTCGCAGATTTTGAATTTGTAGATTGCCAGCAGTTGATCAGTACAGACACATGAT 3977  
Qy 3965 AAGGCTCAACCAACAGAAATGACCTTTCAATTCAGATGATACCTGCTCTGCA 4024  
Db 3978 AAAGCACAACCCAGCAAAATTTGACACACTTCAATCCAAACGGAGTACCCGCGCTTGCAG 4037  
Qy 4025 ACAATTTGCTACCATTTCAAGCAGATCAACGGTATTTTGGCCGGTTTCTCA---GAGCTT 4081  
Db 4038 ACGATTGTGTACCATTTCAAAAGATCAATGCAATATTTGGCCGGTTGTTAGTGAGCTT 4097  
Qy 4082 ACAAGTTGCTGCTCGAGCATTTGATTTCAAGAAAGTTTCTTTTCTTACTAGGAAACT 4141  
Db 4098 ACTAGGCAATTAATCGACAGTGTGATTCGACAGATTTTGTGTTTTCACAAAGAAACA 4157  
Qy 4142 CCAGAAACAGATTTCAAGAAATTTTCTCGGATCTCGACTCGCAGTTCTTATGATGTGTTA 4201  
Db 4158 CCAGCGCAGATTGAGGATTTCTTCGGAGATCTCGACAGTCTGTCGCGATGATGATCTT 4217  
Qy 4202 GAATCTGGATATTTCTAAGTATGATTAAGTACAGAAACAGATTTCAATTGCTGTAGATAT 4261  
Db 4218 GAGCTGGATATTAACAAATACGCAAAATCTCAGAAATGAAATCCACTGTGAGTAGAATAC 4277  
Qy 4262 GAAATATGAAAGATTTGGGCTCTCAATGAGTTTTTGGCGGAGTGTGAAACAAGGGCAC 4321  
Db 4278 GAGATCTGGGAGATTTGGTTTTGAAGACTTCTTTGGGAGAGTTTGGAAACAAGGCAT 4337  
Qy 4322 AGGAAACAACTTTGAGGATTTACATTTGCTGGAATCAAGACATGCTGTGGTATCAAAGG 4381

Db 4338 AGAAAGACCCCTCAAGGATTATACCCAGGTATATAAAACCTTGCACTCTGGTATCAAAGA 4397  
Qy 4382 AAAACGGGTGATGACTACTTTTCATCGGCAATACTGTGTATAAATAGACAGCTTGTCTGGT 4441  
Db 4398 AAGACGGGGACGTCACGACGTTTCAATGGAAACACTGTGATCATTTGCTGTCATTTGGCC 4457  
Qy 4442 TCAATGTTACCGATGGAAAGGTCATAAAAGGTCCTTTTCTGGAGAGAGATTCCTGTTTG 4501  
Db 4458 TCGATGCTTCGATGGAGAAATAATCAAGAGGCTTTTTCGGTGACGATAGTCTGCTG 4517  
Qy 4502 TATTTTCCAAAGGTTTGGATTTCCCTGACATTCAGTCAATGCTGTAAATCTCATGTGGAA 4561  
Db 4518 TACTTTCCAAAGGTTTGGAGTTTCGGATGTGCAACACTCCGGAATCTTATGTGGAA 4577  
Qy 4562 TTTGAGGCCAACTGTATAGAAAGGTPACGGTTACTTTTGTGGTAGATACATCATACAC 4621  
Db 4578 TTTGAAGCAAACTGTTTAAAAAACAAGTATGATACCTTTTCGGGAAGATATGTAATACAT 4637  
Qy 4622 CATGTAAGGAGCAATAGTGTATTTATGATCCTTTGAAGTTGATCTCCAAACTTCGGGCA 4681  
Db 4638 CACGACAGAGATGCAATTTGTGTATTACGATCCCTTAAAGTTGATCTCGAACTTGGTCT 4697  
Qy 4682 AAACATATCAAGGATTTATGATCACTTAGAAGATTAAAGGTTCTTTTGTGCGATGTTCT 4741  
Db 4698 AAACATATCAAGGATTTGGAAACACTTTGAGGAGTTTCAAGGTTCTTTTGTGATGTTCT 4757  
Qy 4742 TGTTGCTCGGAACTGGTGTCTTAGGCTTTCGGAGCTGAAACGACGATTAACAAGAGTT 4801  
Db 4758 GTTTTCGTT---GAACAATTTGCGTATTACACACAGTTGGACGACGCTGTATGGGAGTT 4814  
Qy 4802 CATAAACCGGATGATGTTTGGTTTCTTTTAAATTTGTTGTTTAAACAATTTTGTGAT 4861  
Db 4815 CATAAAGCCGCTTCCAGGTTGCTGTTTATTAAGTCTGGTGAAGTATTTTGTGAT 4874  
Qy 4862 AAATTTTATTTAGAACTTTGTTTAAATGGCTGTTAGTCTCAGAGATACCTGTCAAAAT 4921  
Db 4875 AAAGTTCTTTTATAGATTTGTTTATAGATGGCTCTAGTTGTTAAAGAAAGTGAATAT 4934  
Qy 4922 TAGCGAGTTCAATGATCTTTCCAAACAGGATGAGATACTTCGGCAATTCATGACTAAGT 4981  
Db 4935 CAATGAGTTTATCGACCTGACAAATAATGGAGAGATCTTACCGTCAATGTTTACCCCTGT 4994  
Qy 4982 CAAGAGTTTATAGATATCGACTGTGGACAAAGTATATGCTGTTAAGATGATAGTCTTTC 5041  
Db 4995 AAAGAGTTTATGTTTCCAAAGTTGATAAAATAATGGTTTCAATGAGATGAGTCAATGTC 5054  
Qy 5042 TGATGTAGATTTACTTAAAGGTTTAAAGTATTAGTAAAGAAAGGATGTTGCTGCTAGCTGA 5101  
Db 5055 AGAGTGAACTTTCTTAAAGGATTAAAGTATTGATAGTGGATAGCTCTGTTTACCGG 5114  
Qy 5102 TTTGCTAGTGTCTGGGAGTGGAAATCTCCCGATTAACCTGCGTGGTGGTCAAGTGTG 5161  
Db 5115 TTTGCTGCTACGGCGAGTGGAACTTCCCTGACAAATTCAGAGAGGTTGTGAGCGTGTG 5174  
Qy 5162 TATTTGATTAAGAGATGAAAGGATGAGGAGCAACGCTGGGTGGTATCAGCGCCC 5221  
Db 5175 TCTGTGTGACAAAAGGATGGAAGAGCGGACGAGGCCACTCTCGGATCTTACTACACAGC 5234  
Qy 5222 TGCTTGCAAAAGAAATTTTCTTTTAACTAATCCCTAATTTTCAATAAATCCGAGGA 5281  
Db 5235 AGCTGCAAGAAAGAAATTTTCAAGTCTGTTCCCAATTTATGCTATTAACCAACCAAGA 5294  
Qy 5282 TGCTGAGAGCAACCGCTGGCAAGTGTAGTGAATATCAAGAGAGTGGCTATGGAAGAAG 5341  
Db 5295 CGCGATGAAACCGTCTGCAAGTTTATAGTAAATAGAAATGTGAAGATGTGACGCGG 5354  
Qy 5342 ATACTGCTCTTATCTTTGGAGTTCGTTTCAATTTTGTGTAGTACATATAAATAATGTAG 5401  
Db 5355 TTTCTGCTCGCTTTCTCTGGAGTTTGTGCTGGTGTGATTTGTTTATAGAAATAATATAA 5414  
Qy 5402 AAAAGGTTTGAAGGAAACGATTTTGTGAGTGTGACAGCGCTCGCCAAATTTGACTCACTGA 5461  
Db 5415 ATTAGTTTGAAGAGAAAGATTTACAAACGTGAGAGCGGAGGCCCTTACAGGA 5474

```
QY 5462 AAAGGTGTTGAGGAGTTCGTGCGATCAAGTACCAATGGCTGTGAACCTCGAAAGGTTTC- 5520
D 5475 AGAAGTCGTGTGATGAGTTCATGCGAGAGTCCCTATGTCGATCAGCGTTCGAAAGTTTCG 5534
QY 5521 -----CGGAAAAACAAAAAGAAATAGTGTAGGTAATAATGTTAATAAT- 5561
D 5535 ATCTCGAACCGGAAAAAGAGTGTGTCGCAAGGGAAAAATAGTAGTAATGATCGGTTC 5594
QY 5562 -----HAGAAATAATAACAGTGTGTAAGAGGGTTTAAATTTAGGAATTTAGGA 5614
D 5595 AGTCCGCAACAAGAACTAGAAATGTTAAGGAATTTGGAGGAATGAGTTTAAAAAGAA 5654
QY 5615 TAATGTAAGTATGATGACGAGT-----CTATCGCGTCATCGAGTACGTTTTTAATCAAT 5665
D 5655 TAATTTAATCGATGATGATTCGGAGGCTACTGTCCCGGATCGGATTCGTTTA---AAT 5711
QY 5666 ATGCTTATACAACTCACTCCGAGCCAAATTTGTTTACTTATCTTCGCTTACAGAT 5725
D 5712 ATGCTTACAGTATCACTACTCACTCATCTCAGTTCTGTTCTTGTATCAGCGTGGCCGAC 5771
QY 5726 CCGTGTGAGCTGATCAATCTGTGTAACAAATGATGGGTAAACAGTTCACACGCAACAA 5785
D 5772 CCAATAGAGTTAATTAATTTATGTAATGCTTATAGGAATCAGTTTCAACACAAACAA 5831
QY 5786 GCTAGGACAAACAGTCCAAACAGCAATTTGCGGATGCTCGAAACCTGTGCTAGTATGACA 5845
D 5832 GCTCGAATGCTGTTCAAGACAAATTCAGTGAGGTGTGGAACCTTACCACAACTAAT 5891
QY 5846 GTGAGATTCCTGATCGGATTCCTATGTTATAGATATAATTCGACGCTTGATCCGTTG 5905
D 5892 GTTAGGTTCCCTGACAGTGAATTTAAGGTGTACAGTACAAATGCGGTATTAGACCCGCTA 5951
QY 5906 ATCAGCGGCTTAAATAGCTTTGATAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAG 5965
D 5952 GTACAGCACTGTTAGGTGCAATTCGACACTAGAAATAGAAATAGAAATAGAAATAGAA 6011
QY 5966 CCGCACCGAATPACTACTGAAATCGTTAACGCGACTCAGAGGGTAGACGATCCTACTGTA 6025
D 6012 GCGNACCCACGACTGCCGAACGTTAGATGCTACTCGTAGATGAGACGACGCAAGTGT 6071
QY 6026 GCTATAAGGCTTCAATCAATTAATTTGGCTTAATGAACTGGTTGCGGAACTGGCAATGTC 6085
D 6072 GCCATAAGAGCGCGATATAATTAATTAATGATGAGAAATGATCAGAGGAACCGGATCTAT 6131
QY 6086 AATCAGCAGGCTTTGAGACTCTAGTGTGACTGTGACCACTGTCGACCACTCCGCTACTAG 6145
D 6132 AATCGAGGCTTTTCGAGAGCTCTCTGTTTGGTTTGGACCTCTGGTCTCTGCAACT--- 6188
QY 6146 CTATTTGTTGTGAGATTTCTTAAATAAAGTCCGCTGAAGACTTTAAATTCAGGGTGGCTGA 6205
D 6189 -TGAGGTAGTCAAGATGCATATAATAAATACGGAATTTGTCCGTAATCACA-CGTGTCGG 6246
QY 6206 THACAAATACAGAGTGGTTGTTGTCGTCATTAATAATATAACAAATGTCATATCTGGATCC 6265
D 6247 TACGATAAGCGATAGTGTGTTTCCCTCCCACTTAATTCGAAGGTTGT-GTCTTGTGATCGC 6305
QY 6266 AACAGTTAAACCATGTATGTTGTTATCTGTGTTATGCGTAAACATCGAGAGGTTTCG 6325
D 6306 GCGGTCAATATGTTATGTTTCATATACATCCGAGGCAACGTAATAAAGCGAGGGGTTTCG 6365
QY 6326 AATCTCCCTAACCGCGGTAGCGGCCCA 6355
D 6366 AATCCCGCGTTTACCCCGGTAGGCGGCCA 6395
```

## RESULT 9

AAZ20643

AAZ20643 standard; RNA; 6439 BP.

XX

AC AAZ20643;

XX

DT 26-NOV-1999 (first entry)

```
XX TMV-based virus TMV261 coat protein read-through RNA sequence.
D 16-SEP-1999.
KW TMV-based virus; tobacco mosaic virus; protein isolation; green juice;
X virus isolation; fusion protein identification; ss.
X Tobacco mosaic virus.
X WO9946288-A2.
X 09-MAR-1999; 99WO-US005056.
X 10-MAR-1998; 98US-00037751.
X (BIOS-) BIOSOURCE TECHNOLOGIES INC.
X Garger SJ, Holtz RB, McCulloch MJ, Turpen TH;
X WPI; 1999-561660/47.
X Obtaining protein, viruses and fusion proteins from plants, using non-
X denaturing conditions.
X Disclosure; Page 43-47; 58pp; English.
X This sequence represents a tobacco mosaic virus (TMV) based virus
X sequence identified using the method of the invention. The method is for
X obtaining a soluble protein or peptide of interest from a plant,
X comprises homogenising the plant to produce green juice, adjusting the pH
X to less than or equal to 5.2, and heating the juice to a minimum of 45
X degrees C. The juice is then centrifuged to produce a supernatant, and
X the protein or peptide is purified from the supernatant. The method can
X also be used for obtaining viruses and fusion proteins. The method is
X especially useful for obtaining IL-1 to IL-10, EPO, G-CSF, GM-CSF, hp-
X CSF, M-CSF, Factor VIII, Factor IX, tPA, receptors, receptor antagonists,
X antibodies, vaccines, single-chain antibodies, enzymes, neuropeptides, insulin,
X antigens, hormones, peptide hormones, calcitonin, and human growth
X hormone, or an antimicrobial peptide or protein from proteoglycans,
X magainins, ceropins, melittins, indolicidins, defensins, beta-defensins,
X cryptidins, clavalins, plant defensins, nicot and batenecins, all
X produced by recombinant means. The new method is more efficient than the
X prior art for isolating viruses, protein, and peptides. The method is
X large-scale, and non-denaturing and solvent-limited. Prior art methods do
X not isolate recombinant proteins, and do not allow fraction 2 proteins to
X be ultrafiltrated
X SQ Sequence 6439 BP; 1873 A; 1240 C; 1555 G; 0 T; 1771 U; 0 Other;
X Query Match 38.6%; Score 2455.4; DB 2; Length 6439;
X Best Local Similarity 44.8%; Pred. No. 0;
X Matches 2772; Conservative 1147; Mismatches 2206; Indels 60; Gaps 8;
QY 20 ACAACAAATTAACAAACAAACAAATTTACAAACAAACAAACAAACAAACAAACAAAC 79
D 18 ACCAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 77
QY 80 ATACAAATTAATTAAGCAACGCCCTTTTGAAGCGTGTGTTAAACAACTTCGTT 139
D 78 ACACAGACAGTACCAACACAGCUUUGCUGACACUGCCGAGGAAACAAACUCCUUGGUC 137
QY 140 AATGACCTTCGAAGGCGCATGTACGATACGCGCGTGGAGAAATTTAAACGCCCGCAC 199
D 138 AAUGAUCUAGCAAGCGUGUCUUAACACACACGCGGUGAAGAUUAACGUCUGAC 197
QY 200 CGTAGACCAAGGTCACCTTTTCCAAAATATTAGCGAAGACAAACGCTTCTAGTCTCC 259
D 198 CGCAGGCCCAAGGUGAACUUUUCAAAAGAAUAAAGCGAGGAGCAGACGCUUAUUGCUACC 257
QY 260 AACGGTACCGGAGTTCAGATTACCTTTTATATCTTATCTCAAAATGCGGTACACAGTTTG 319
D 258 CGGCGUAUCCAGAAUUCAAAUAACUUUAUUAACACGCAAAAUUGCCGCAUUCGCUU 317
```

Qy	320	GCTGGAGTTTGAGAGCATTTAGAATTGGAATATCTGATGCTACAAGTTCCTCTATGATCG	379
Dd	318	GCAGGUGAUAUUCGUAAACUUGAUAUUCUGAUGAUGCAAAUUCUCCUACGCAUCA	377
Qy	380	CCGACATATGATATAGTGTGGAACTTTGCGAGCACATTTGTTCAAAGCAGGGAATTACGTG	439
Dd	378	UUGAUUUAGACAUAUGCGGGAUUUUUGCAUCCGACUCUCUUAAGGACAGACAUUGUA	437
Qy	440	CATTGCTGATGCCCAATCTGGACATACGAGATATATAGGACGACGAAGACAAAAAGGAC	499
Dd	438	CACUGUGCAUGCCCAACCUUGHCGUUCGHAGACAUC AUGCGGACGHAAGCCGHAAGAC	497
Qy	500	TCAATTGAGATGTAATTTGTCGAGATTGTCTCGTTCTAA CAAGGTAA T TCCTGAGTTTCAA	559
Dd	498	AGUAUUGAACUAUAUACCUUUCVAGGCUAGAGAGAGGGGGGAAAAACAGUCCCAACUCCAA	557
Qy	560	AGGAGGCTTTTTAA CAGGTATGCAGAAGCTCCCA CGAAGTCTGCTGCTCTAAACCTTTT	619
Dd	558	AAGGAAGCAUUGAGACAGCAUAUCGAGAAUUCUCUGAAGACGUGUGUCUCAUAUACUUUC	617
Qy	620	CAGGATTGTCGAATACATCCGCCAGAGAA TAGTGGTAGAAGATACGCTGTGCTCTGCAC	679
Dd	618	CAGACAUGCGACAUCAGAGCCGAUCGACAAUCAGGCHAGAGUAGUCCAUUUGCGUACAC	677
Qy	680	AGTTTGTATGATA TTCTGTGTCATGAGTTTGGAGCTGCGTTAA TATCTAAGAATATACAT	739
Dd	678	AGCAUAUAUGACAUAACCAGCGAGAUUCGGGGCGGCAUCUUGAGGAAAAAUGUCCAU	737
Qy	740	GTAATGTTAGCAGTTCATTTTGGCAGAGCATATTATTA CTAGACCHAGCGGAGGTTACG	799
Dd	738	ACGUGCUAUGCGGUUUCACUCCUCGAGAA CCUGUCUCUGAAGAAUUAUACUGUCAU	797
Qy	800	CTTAAAGAAATAGGCGCAACTTTTCAAAGAGAGGATGATGTTTCTTTTCTTTGCT	859
Dd	798	UUGGACGAAAUCAACCGGUGUUUUUCCGGCGAUGGAGACAAGUUGACCUUUUUCUUUGCA	857
Qy	860	GATGAAAGTACTTTTAATATTAGTCAATAAATACAAAATATCTTGCATTATTAGTTAAA	919
Dd	858	UCAGAGAGUACUUAUAUAUUGUCAUAUUUAUUCUAUAUUCUUAAGUAUGUGUCAA	917
Qy	920	TCATTACTTTCCTGCTTCTAGTAGAATAGTTTACTTTTAAAGGAA TTTTATGTCATAGGTT	979
Dd	918	ACUUAUUCGCGCCUUAUAUAGAGAGUUUAU A UGAAGAAGUGAUUUUAUGUCACAGAGUU	977
Qy	980	AATACTTGTGTTTGTAAATTTTACCAAGTAGATACCTATATCTGTGTACAGAGGTTAGA	1039
Dd	978	AUAUCCGGUUUUGAAGUUUUCAGAAUAAGUA CUUUUCUUUUGUA CAAGGUGUGGCC	1037
Qy	1040	CAAGTAGGTTGTATGATGTAGTTCAGTTCTATGAGCGCATGGAAGACGCTTTGTCTTACAAG	1099
Dd	1038	CAUAAAGUAGUAUGAGGAGCGAUUUUAUUCGCAUUGNAGACGCAUGGCAUUAACAA	1097
Qy	1100	AAAACCTTGCCCATGTTTCAA CACTGAAAGAGCAATCTTTAGACACACGGCTTCGGTTAAC	1159
Dd	1098	AAGACUCUUGCAUUGUCAAACAGCGAGAGAAUCCUCCUGAGGAUUAUCAUCAUGUCAU	1157
Qy	1160	TTTTTGGTTCCCTAAGATGAAGGACATCGTGATAGTACCGCTGTTTGGGGTTCTATTACC	1219
Dd	1158	UACUGUUUCCCCAAAUDGAGGGAUAUGGUCAUCUGUACCAUUAUUCGACA UUUCUUUGGAG	1217
Qy	1220	AGCAAAAGATGACAAGGAGTGAGTTCATTTGTTTAA TCGTGACTTCGTTTACACAGTGCCTT	1279
Dd	1218	ACUNGAAGAGGACCGCCAAGGAAGUUCUUAUGUGUCGAAGGAUUCUGUGUUAUACUGUCU	1277
Qy	1280	AATCATATCAGAACATATCAAGCCAAAAGCGTTAACTTTACCAGAACGPTATTATCTTTGCTG	1339
Dd	1278	AAACCACAUUCGAA CAUACCGGCGAAAGCUCUUAUCAUAUACGAAAUUUUUGCCUUGUC	1337
Qy	1340	GAGTCTATAGATCCC GGTGTAATCAATG GTGTGTTACTGCTAGGTCGTAATGGGATGTA	1399
Dd	1338	GAUUCGAUUCGAUCGAGGGUAAUCAUUAACGGUGUGACAGCGAGGUCCGGAUUGGGAUGUG	1397



















Db 3558 GUUACCAUGAGUUGACUGACAUUUCUAUUGAAUGUCAAAGAUUGCAUUAUGGAUUGUCU 3617  
Qy 3605 AAAAGATATTCGATGTCACCAAGAGAGTGAAACCATGCTCTAGAGCCAGTTTTCGCTACCGCG 3664  
Db 3618 AAGUCUGUUGUCGCGCCUAGGAUCAAUCAAACCAUUAUACCUUUGUACGAACGCGC 3677  
Qy 3665 CGGNAACCGCCAGCGGCTGAGGACTACTCGAAATCTGGTTGCAATGATTAAGAAAT 3724  
Db 3678 GCAGAAUUGCCACGCGACAGUCGUAUUGGAAAUUUAUGGCGGCGAUGAUUAAAGGAAC 3737  
Qy 3725 TTCACGACACAGACCTCAGCGGGAGGATTCGACTTGAGAGACACGCGCATCTGTTGTAGTA 3784  
Db 3738 UUUAAACGCGCGAGUUGUCUGGCAUCAUGAUUUUAGAAUAUACUGCAUCUUUAGUUGUA 3797  
Qy 3785 GATAAGTTTGTGATGATTTTATTAAGAAAGAAATACAAAAAATATTGCTGGA 3844  
Db 3798 GAUAGUUUUUGAUUGAUUUUGCUUAAAGAAAAAGAAACCAAAUAAAUUUGUUCU 3857  
Qy 3845 GTGATGACGAAGGATTCGAATGATGATGTTGGAAAAACAGAAAGAAAGTACTATTGGAC 3904  
Db 3858 UUGUUCAGUAGAGAGUCUCUCAAUAGUUGUUAAGAAAGACGAAACAGGUAAACAAUAGGC 3917  
Qy 3905 GACTTGGCTAACTACAAATTTTACAGATCTGCGGCCATCGATCAGTACAAGCAATGATC 3964  
Db 3918 CAGCUCGAGAUUUUGAUUUUGUAGUUUGCCAGCAGUUGUACGUAACAGACAUAGAU 3977  
Qy 3965 AAGGCTCAACCAAAAACAGAAATGACCTTTCAATTCAGAATGAATACCTGCTCTGCA 4024  
Db 3978 AAAGCAAAACCCAGCAAAAUUGGACACUCUCAAUCCAAACGGAGUACCCGCUUGCAG 4037  
Qy 4025 ACAATGCTCTACCAATTCGAAGCAGATCAACGGTATTTTGGCGGTTCT---CAGAGTT 4081  
Db 4038 ACGAUUGUGUACCAUCAAAGAAAGAUCAUUGCAUUAUUGGCCGCGUUGUAGUGAGCU 4097  
Qy 4082 ACNAGTTGCTGCTCGAGGCAATTTGATTTCTAAGAGTTTCTTTTCTTACTAGGAAACT 4141  
Db 4098 ACUAGGCAAUUACUGGACAGUGUUGAUUCGAGCAGAUUUUUUUUUCACAAAGAAAGCA 4157  
Qy 4142 CCAGAAACAGATTCAGAAATTTTCTCGGATCTCGACTCGCAGTTCCTCTATGGATGTGTA 4201  
Db 4158 CCAGCGCAGUUGAGGAUUUCUCCGAGAUUCUGACAGUCAUGUCCGAGGAGUUGUCUG 4217  
Qy 4202 GAACTGGATATTTCTAAGTATGATTAAGTCAAGAACGAGTTTCAATGTCGTGTAGAGTAT 4261  
Db 4218 CAGCUGGAUUAUCAAUAAUACGACAAUUCAGAAUAGAAUCCACUGUGCAGUAGAAUAC 4277  
Qy 4262 GAAATATGGAAGAGTTGGTCTCAATGATTTTGGCGGAGTGTGGAACAAAGGCAC 4321  
Db 4278 GAGAUUCGGCGGAAGAUUGGGUUUUGAAGACUUCUUGGAGAGAUUUUGAAACAAAGGCAU 4337  
Qy 4322 AGGAAACCAACTTTGAAGGATTCATCTGGAATCAAGACATGTCGTGGTATCAAAGG 4381  
Db 4338 AGAAAGACCACCUCAAAGAUUAUACCGCAGGUAAUAAACUUGCAUCUGGUUAUCAAAGA 4397  
Qy 4382 AAAAGCGTGATGTGACTTTCTATTCGCGCAATCTGTTATAATAGCAGCTTGTGGGT 4441  
Db 4398 AAGACGGGACGUCACGACGUCAUUGGAAACACUGUGAACACUGUGCAUGUUUGGCC 4457  
Qy 4442 TCNATGTTACGATCGAAAGGTCAAAAGTGCTTTTGTGGAGACGATTCGTTTTG 4501  
Db 4458 UCGAUGUUCGAGGAGAAAUUAUCAAAGGAGCCUUUUGCGGAGCAGUAGUCUGCUG 4517  
Qy 4502 TATTTTCAAGGGTTTGATTTCCCTGACATTCAGTATGTCGTAATCTCATGTGGAAT 4561  
Db 4518 UACUUUCAAAGGGUUGUGAUUUCGGAUGUGCAACUCCGCGCAUUCUUUUGGAAU 4577  
Qy 4562 TTTGAGGCCAACTGTATAGAAAGGTAACGTTACTTTTGTGGTAGATACATATACAC 4621  
Db 4578 UUGAAGCAAAACUGUUUAAAAACAGUAGUAGUACUUUUGGGAAGAUUAGUAAUACA 4637  
Qy 4622 CATGATAGGGAGCAATAGTATTATGATCTCTTTGAAGTTGATCTCCAAACTTGGGCA 4681

Db 4638 CACGACAGAGUACUUGUUAUUAUACGAUCCCTUAAGUUGAUUCUGAAACUUGUGCU 4697  
Qy 4682 AAACATATCAAGGATTAATGATCACTTAAAGAGTAAAGGGTGTCTTTGTGCGATGTTGCT 4741  
Db 4698 AAACACAUCAAGAUUGGAAACACUUGAGGAGUUCAGAAAGUCUCUUGUGAUGUUGCU 4757  
Qy 4742 TGTTCGCTCGGAACACTGGTCTTAGGCTTTCCGCGAGCTGAACGCGCTATCAAGGAGGTT 4801  
Db 4758 GUUUCGUUGAACAAUUGUGCGUA---UUAACAACAGUUUGGACGCGUUGAUGGGAGGU 4814  
Qy 4802 CATAAAAACCGCATTTGATGTTTCGTTTCTTTTAAATTTGTTTAAACAAATTTTGTGTAT 4861  
Db 4815 CAUAAAGACCGCCUCCAGGUUCUUUUUAUAAAGUCUGGUGAGUAUUUGUCUGAU 4874  
Qy 4862 AAATTTTATTTAGACTTTGTTTTTAAATGGCTGTTAGTCTCAGAGATACCTGTCAAAAT 4921  
Db 4875 AAAGUCUUUUUAGAAUUGUUAUAGAGGCUUAGUUAAGGAAAGGAAAGUAAU 4934  
Qy 4922 TAGCGAGTTCAATTCATCTTTCGAAACAGGATGAGTACTTCCGGCATTCATGACTTAAGGT 4981  
Db 4935 CAUAGAUUUUAUCCACCUCAAAAAGGAGAGAUUUUACCGUGCAUGUUUACCCUGU 4994  
Qy 4982 CAAGAGTGTAGAAATATCGACTGTGGACAAGATTATGGCTGTTAAGAAATGATAGTCTTTC 5041  
Db 4995 AAAGAGUUAUUGUUCUCAAAGUUGAUAAUAAUUGUUAUGAUGAUGAUCAUUGUC 5054  
Qy 5042 TGATGTAGATTTACTTAAAGGTGTTAAGTTAGTTAAGAAAGGTTATGTTGCTTAGCTGA 5101  
Db 5055 AGAGUGAAACCUUUAUAAAGGAGUUAAGCUUUAUGAUGAUGAUGUUAUGCGUUAUGCGG 5114  
Qy 5102 TTTGTTAGTGTCTGGGAGTGGAACTCTCCCGATACTGCCGTGGTGTGTCAGTGTGTTG 5161  
Db 5115 UUGUGUCGUCACGCGCGAGUGGAAACUUGCCUGACAAUUGCAGAGAGGUGUGAGGUGUG 5174  
Qy 5162 TATTGTAGATAAGAGAAATGAAAGGAGTAAGAAAGCAACGCTGGTGGTATCACGCCCC 5221  
Db 5175 UCUGUGGACAAAGAGUAGAAAGCGCAGCGCCACUUCUGGAUUAUUAUUAUUAUUA 5234  
Qy 5222 TGCTTGCAAAAGAAATTTTCTTTTAAAGCTAATCCCTAATTTATTAATAAATTAATGTAAG 5281  
Db 5235 AGCUGAAAGAAAGAUUUCAGUUCUCAAAGGUGUCCCAAUUAUUGCUAUAACCAACCCAGA 5294  
Qy 5282 TGCTGAGAGACACCGGTGGCAAGTGTAGTGAATCAAGAGAGTGGCTATGGAGAGG 5341  
Db 5295 CGCGAUGAAACGUCUGGCAAGUUUAGUUAUUAUAAUUAUAAUUGAUGAUGAUGUUGACGGG 5354  
Qy 5342 ATACTGCTCTTATCTTTGAGATTCGTTTCAATTTGTTAGTACATAATAATAATGTAAG 5401  
Db 5355 UUCUGUCCGUUUCUGGAGUUUUGUGCGGUGUUAUUGUUAUAGAAUUAUAGAAUUAUAAA 5414  
Qy 5402 AAAAGGTTTGAGGAAACGATTTTGTAGTGTGACAGACGCTCGCCAAATTGAATCACTGGA 5461  
Db 5415 AUUAGGUUAGAGAGAGAAUUAUCAAACGUGAGAGACGAGAGGCCCAUGGAAACUUAACA 5474  
Qy 5462 AAAAGTTTTCAGAGTTCTGGATGAAGTACCAATGGCTGTGAACTCGAAAGGTTGCC 5521  
Db 5475 AGAAGUCGUGAUGAUGUUAUGGAAGAUUGCCUUAUGUCGAGUCAGGCUUGCAAGUUUGC 5534  
Qy 5522 GG-----AAAAAAGAAATGTTAGTAAATTAATGTTAATAAT----- 5561  
Db 5535 AUUCUGAACCCGAAUAAAGAGUAGUUGCCGAAAGGGAUAAUAGUAGUAGUAGUAGUAG 5594  
Qy 5562 -----AAGAAATAATAACAGTGTGTAAGAGGGTTTTTAAATTTAGGAAATTTAGGA 5614  
Db 5595 AGUGCCGAAACAAGAACUAUAGAAAUUUAAGAUUUUUGGAGGAUAGAGUUAUAAAAAGAA 5654  
Qy 5615 TAATGTAGTATGATGACAGT-----CTATCGCTCATCGAGTACGTTTATCAAT 5665  
Db 5655 UAAUUAUUAUCGUAUGAUGAUGCGGAGGCUACUGUGCGCAUUCGGAUUCGUUUA---AAU 5711  
Qy 5666 ATGCTTATACAACTCAACTCTCCGAGCCAAATTTGTTTACTTATCTTCCGCTTTACGCAGAT 5725  
Db 5712 AUGUCUUAAGUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 5771











Db 5235 AGCUGCAAGAAAGAAUUCAGUUCAGGUCGUCUCCAAUUAUGCUAAACCCAGGA 5294  
Qy 5282 TGCTGAGAGCACCGCTGCGAAGTGTAGTGAATATCAAGAGAGTGGCTATGGAAGAGG 5341  
Db 5295 CGCGAUGAAACAGCUGCGCAAGUUAUUAUAAUAAUAGUAGUAGUAGUAGCGGG 5354  
Qy 5342 ATACTGCTCTTATCTTTGGAGTGTCTGTTCAATTTGTGTAGTACATATAAATAATGTAAG 5401  
Db 5355 UUCUGUGCGUUCUUCUGUGAGUUGUGUGUGUGUGUUAUAGAAUAAUAAUAA 5414  
Qy 5402 AAAAGTTTGGAGGAAACGATTTTTCAGTGTGACAGCGCTCGCAATTTGAATCACTGA 5461  
Db 5415 AUUAGUUUAGAGAGAAAGAUUACAAACGAGAGAGCGAGGCGCCAUUGGAAUUAACAGA 5474  
Qy 5462 AAAGTTTGTGAGGAGTTCGTGGATGAAGTATACCAATGCTGTGAACCTCGAAAAGTTTCC 5521  
Db 5475 AGAAGUUGUAGUAGUUAUGGAAGUUGUCCUAUGUGCAUGCAUGGCUUGCAAGUUG 5534  
Qy 5522 GG-----AAAAAAGAAAGTGTAGTGAATTAATTAATTAAT----- 5561  
Db 5535 AUCUGAACCGGAAAGAGUAGUUGCGCAAGAGGAAUUAUAGUAAUAGUAAUAGUAG 5594  
Qy 5562 -----AAGAAATTAATACAGTGTGAAGAGGTTTAAATTTGAGGAAATTTGAGA 5614  
Db 5595 AGUGCCGAAACAGAAACUUAAGAAUUGUUAAGUUAUGGAGAAUAGUUAUAAAGAA 5654  
Qy 5615 TAATTAAGTATGATGATGAGT-----CTATCGCTCATCGATGAGTGTGTTTAATCAAT 5665  
Db 5655 UAAUUAUAGUAGUAGUAGUUGGAGGCUACUGCGCCGAAUUGGAAUUGUUAUAA 5711  
Qy 5666 ATGCTTATACAAATCAATCTCGAGCAAAATTTGTTTATCTATCTTCGCTTACGAGAT 5725  
Db 5712 AUGUUAUAGUUAUAGUUAUAGUUAUAGUUAUAGUUAUAGUUAUAGUUAUAGUUA 5771  
Qy 5726 CCGTGCAGCTGATCAATCTGTGTACAAATGATGCTGATGCTGATGCTGATGATGAT 5785  
Db 5772 CCAAUAGAGUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 5831  
Qy 5786 GCTAGGCAACAGATCCAAACAGCAATTTTCGATGCTGCTGCTGCTGCTGCTGCTGCT 5845  
Db 5832 GCUCGAAUUGUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 5891  
Qy 5846 GTGAGATTTCTGATCGGATTTCTATGTATAGATATATTCGACGCTTGTGCTGCTGCT 5905  
Db 5892 GTUAGGUUCCUGACAGUGACUUAUUAAGUGUACAGGUACAAUUGCGUUAUAGACCGCUA 5951  
Qy 5906 ATCAGCGGTTATTAATAGCTTTGATGATGATGATGATGATGATGATGATGATGAT 5965  
Db 5952 GUCACAGCAGUUAUAGGUGCAUUCGACACUAGAAUUAUUAUUAUUAUUAUUAUUA 6011  
Qy 5966 CCGCAGCAGGATCTACTGAAATCGTTAAACGCGACTCAGAGGTTAGACGATGCTACTGTA 6025  
Db 6012 GCGAACCCACGACUGCGGAAACGUAUAGUAGUUAUUAUUAUUAUUAUUAUUAUUA 6071  
Qy 6026 GCTATAAGGGTCTCAATCAATATTTGCTATGATGATGATGATGATGATGATGATGAT 6085  
Db 6072 GCGAUAAGGAGCGGCAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 6131  
Qy 6086 AATCAAGCAGGCTTTGAGACTGTAGTGTGCTGAGCAGCAATCTCGGCTACTTAG 6145  
Db 6132 AAUCGAGCUCUUCUGAGAGCUCUUCUGGUGUUGUUGUUGUUGUUGUUGUUGUUG 6191  
Qy 6146 CTATT 6150  
Db 6192 CAAUU 6196

RESULT 13  
ID AAZ20646  
XX AAZ20646 standard; RNA; 6446 BP.  
AC AAZ20646;

XX 26-NOV-1999 (first entry)  
DT TMV-based virus TMV861 coat protein read-through RNA sequence.  
XX TMV-based virus; tobacco mosaic virus; protein isolation; green juice;  
DE virus isolation; fusion protein identification; 88.  
KW Tobacco mosaic virus.  
KW W09946288-A2.  
XX 16-SEP-1999.  
XX 09-MAR-1999; 99WO-US005056.  
XX 10-MAR-1998; 98US-00037751.  
PA (BIOS-) BIOSOURCE TECHNOLOGIES INC.  
XX Garger SJ, Holtz RB, McCulloch MJ, Turpen TH;  
XX WPI; 1999-561660/47.  
XX Obtaining protein, viruses and fusion proteins from plants, using non-  
PT denaturing conditions.  
XX Disclosure; Page 55-58; 58pp; English.  
XX This sequence represents a tobacco mosaic virus (TMV) based virus  
CC sequence identified using the method of the invention. The method is for  
CC obtaining a soluble protein or peptide of interest from a plant,  
CC comprises homogenising the plant to produce green juice, adjusting the pH  
CC to less than or equal to 5.2, and heating the juice to a minimum of 45  
CC degrees C. The juice is then centrifuged to produce a supernatant, and  
CC the protein or peptide is purified from the supernatant. The method can  
CC also be used for obtaining viruses and fusion proteins. The method is  
CC especially useful for obtaining IL-1 to IL-10, EPO, G-CSF, GM-CSF, HP-  
CC CSF, M-CSF, Factor VIII, Factor IX, tPA, receptors, receptor antagonists,  
CC antibodies, single-chain antibodies, enzymes, neuropolypeptides, insulin,  
CC antigens, vaccines, peptide hormones, calcitonin, and human growth  
CC hormone, or an antimicrobial peptide or protein from proteoglycans,  
CC magainins, cecropins, melittins, indolicidins, defensins, beta-defensins,  
CC cryptidins, clavinins, plant defensins, nicin and bactericins, all  
CC produced by recombinant means. The new method is more efficient than the  
CC prior art for isolating viruses, protein, and peptides. The method is  
CC large-scale, and non-denaturing and solvent-limited. Prior art methods do  
CC not isolate recombinant proteins, and do not allow fraction 2 proteins to  
CC be ultrafiltrated  
SQ Sequence 6446 BP; 1873 A; 1234 C; 1563 G; 0 T; 1776 U; 0 Other;  
Query Match 38.5%; Score 2449.8; DB 2; Length 6446;  
Best Local Similarity 44.8%; Pred. No. 0;  
Matches 2771; Conservative 1146; Mismatches 2212; Indels 60; Gaps 8;  
Qy 20 ACACACAAATTAATAAACAACAATATTAACAACAACAACAACAACAACAACAACA 79  
Db 18 ACCAACACACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 77  
Qy 80 ATACATCTATATTAATAGCAACGCTTCTTGAAGCGTGTGTTAAACAACAACAACA 139  
Db 78 ACACAGACAGCAUACCAUACGUCGUGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 137  
Qy 140 AATGACCTTTCAGAGAGGCGCATGTACGATACGCGCGTGGAGGAGGAGGAGGAGG 199  
Db 138 AAUAGUUAAGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 197  
Qy 200 CGTAGACAAAGGTCAACTTTTCCAAAACCTATTAGCGAAGAGCAAAACGCTTCTAGTCTCC 259  
Db 198 CGCAGGCCCAAGGUGAACUUUUAAGAAUUAAGGAGGAGGAGGAGGAGGAGGAGGAGG 257  
Qy 260 AACGGTACCCGGAGTTCAGATTACCTTTTATTAATCTCAAAATGCCGTACACAGTTTG 319



Qy	2474	TCGTGATATTCGAAGCTCCAAAATCTCGAGAAAACAATGTAGAGACGGTGTGAACCCCAAGAA	2533
Db	2478	UCCGACAUAGCGCGAAAUCUCAGAAUCUCGCGCAGACUGUCUGGAAAACGAGAACCGCAUGUC	2537
Qy	2534	CCTACTCGAAGATGGTACTTGTGTGATGGGGTGCTGTGTTGTGGAAGTACAAAGGAGAT	2593
Db	2538	AGUAGCCGAAGGUGUUGUUGUGGACGGAGUUCCGGGCUGUGGGAAAAACAAGAAAUU	2597
Qy	2594	TTTGAAGAATTTGATCTTTGATGAGGATTTGATCTTTGTTCTCTGCGAAAACAAGCTGTGCT	2653
Db	2598	CUUCCAGGGUUAUUUGAUGAAGAUUAUUUAGUACUCUGGGAAGCAAGCGCGAA	2657
Qy	2654	ATGATCAGAAGAGGGCTTAATTCATCTGGACTGATTAAGAGCCACAATGTGGACAATGTGAGA	2713
Db	2658	AUGAUCAGNAGACGUGCGNAUUCUCACGGGAUUAUUGUGGCCACGAAGGACACGUAUAA	2717
Qy	2714	ACGGTAGATTCACTTCTAATGTCATCCAAAACCGCGATCAC-----ACAAGAGGCTT	2764
Db	2718	ACCGUAGUUCUUCAUGAUAUUUGGGAAGAAGCACACGUGUCAGUUCUACAGAGGUUA	2777
Qy	2765	TTTATTGATGAAGGTTTGATCTCTGTCACACACCGTTGTGTTAACCTTCCTGGTGCTTATCTCT	2824
Db	2778	UTUCAUUGAAGGGUUGAUGUUGCAUACUGUUGUGUUAUUUUUUGUGCGAUGUCA	2837
Qy	2825	GGTTGCCACATCGCATACATTTACGGAGATACACAGCAGATTCCTTTCAATTAACAGATT	2884
Db	2838	UUGUGCGAAAUUGCAUAUUGUUAACGGAGACACACAGCAGAGAUUCCAUAACAUAUAGAUU	2897
Qy	2885	CAGAAATTTCCGATCCCAAACATTTTGGAGAAGCTGCAAGTGGATGAAGTTTGAGATGAGG	2944
Db	2898	UCAGGAUUCGCGUACCCCGCCAUUUGCCAAUUGGNAUGUAGCGAGGUGGAGACACGC	2957
Qy	2945	AGGACCACTGAGATCCCGAGTGATGTGAATTTTTTCTCAATCGAAGTACGAAGGA	3004
Db	2958	AGAAUCUACUCUCGUGUCCAGCGAUGUACACAUAUUGAACAAGAGAUUAGAGGC	3017
Qy	3005	CGGTGACACCACTTCAACTGTGTACAGCATCGGTCTCATCTGAGATGATAGCGGGTAAG	3064
Db	3018	UUUGUCAUGAGCACUUCUCUGGUUAAAAGUCUGUUCGCGAGGAGUUGUGCGCGAGCC	3077
Qy	3065	GGAGTACTAAACAGTGTTCCAAACCACTAAAAGGGAAAAATTGTAACTTTCACTCAGGCT	3124
Db	3078	GCCGUGAUCAAUCCGAUCUCAAAACCCUUGCAUGGCAAGAUCCUGACUUAUCCCAUUG	3137
Qy	3125	GATAAATTTGATGTAGAGAGAGAGGGCTATAAGATGTGAAACACCGTTTCATGTAGATCCAA	3184
Db	3138	GAUAAAAGACUCUCUGUUAACAAGAGGGUUAUUCAGAUUACACUGUGCAUGAAGUGCAA	3197
Qy	3185	GGAGAAACCTTTGAAGATGTGTCGCTGTCAGATTGACGCGCACTCCACATGACTCTGATT	3244
Db	3198	GCGGAGACAUACUCUGAUGUUAACUAGUUGGUUAAACCCUACACAGUCUCCAUCAU	3257
Qy	3245	TCCAAGTCTTCCCGCATGTTCTAGTCGCTCTGACTAGACACACAAAGAGCTTCAAAAT	3304
Db	3258	GCAGGACAGGCCACAUGUUGUGCAUUGUCAAGGCACACCUUUGCUCUACGUAC	3317
Qy	3305	TACACCGTAGTGTAGATCTCTTTTAGTACAGATAATTAGTATTTGTCTCTTTAAGCTCC	3364
Db	3318	UACACUGUGUUAUGGAUCCUUUAGUUAUACAUAUAGAUUACUAGAGAAACUUAUGCUG	3377
Qy	3365	TTCCTTTTGAAAATGTATATGGTAGAAGCAGGTAGTAGATAGCAATTTACAGATGGATGCA	3424
Db	3378	UACUUGUUAUUAUAGUUAAGUCUAGCAGAGAAACAAUAGCNAUUAACAGAUUACUUG	3437
Qy	3425	GTGTTCAAAGGTCAATAATCTCTTTGTGGCAACACCTTAAATCAGGAGACTTTTCCAGATCTA	3484
Db	3438	GUGUUCAAAGGUAUUAUUUUGUGCAGCGCCAAAGACUGGUGAUUAUUUCUGAUUG	3497
Qy	3485	CAGTTCATTATAGATGATGCTCCCTCGGTATAATAGTACTATATTAACAAGATATGATCT	3544
Db	3498	CAGUUUAUUAUUAUAGUUGUCUCCCGAGCAACAGCCAAUGAUGAUAUUAUUUGAUGCU	3557







Db 1938 CAAGAGAGGCUUCAGAGGUGUUCUUGGUAUUCUUCAGAGAGAGUUGAAGAACCGUCC 1997  
Qy 1994 AATAAATTTTCTATTGCTGAGAAAGGGAGATTGCTGTGTGTCGAGAAAGTCTATGTTTG 2053  
Db 1998 AUGAAGGUGUUGAUGCCAGAGAGAGUUAUUAUUGAGUGGUGUUCUGGAGUAUCUGG 2057  
Qy 2054 ACGAATGCTAACTTAGAGCACCAGGAGTTGGAGTCCCTCAACGATTTCCATAGGCTTGC 2113  
Db 2058 GAGUGUCUUAUUCUUAAGAACGAGAGAUAGAGUCUUAUAGACGAGUUAUUAUAGCAAG 2117  
Qy 2114 GTGATAGTGTGATTAACAAGCAATGGCATCGGTTGTCTACATGGCTCACTCAAGTT 2173  
Db 2118 GCGAUUCUGUUAUUGAAGCAGAGAGUAGUCUGAUCAACCGGUCGGAUUAAGUU 2177  
Qy 2174 CAACAAATGAAGAACTATGTGACAGTTTGGCAGCTTGTCTCGCCACTCTATCAAT 2233  
Db 2178 CAGCAAAUAGAAACUUUAUCAUAGCUGGUAUCAUCAUUGUGCGGUGUGGAU 2237  
Qy 2234 CTATGCAAGTCACTAAGAGATGAAGTCCGGTATGATTTCTGATTCAGGGAGAAAGTTGT 2293  
Db 2238 CUCGUCAAGUCCUCAAAGAUACAGCUGCUUAUUGACCUUUAACCGUCAAAGUUGGA 2297  
Qy 2294 GTTGGGATGTCATTTGAAAGTGGCTCTCAACCTGGCGGCAAGGTCAATTCATGG 2353  
Db 2298 GUCUGGAGUGUUGCAUCUAGGAAGUGUUAUUAUCAAACCAACCGGCAAGAGUAGUAGUGG 2357  
Qy 2354 GAGATTGCTCTGATTTACAAAGGGGAAATGTTTACTGCACTTCTATCTTATCAAGAGAT 2413  
Db 2358 GUGUUGUUGAAACCCACCGAGGAAGUAUCAUGUGCGCUUUGGAUAUAGUAGCAG 2417  
Qy 2414 AGAATGGTGAAGTGAAGCGATGGAGAGGGTGTATCTATCTATCAATAGTATAT 2473  
Db 2418 GUGUGGUGACAUUGCAUGAUUGGAGAGAGUAGUGUGCAGCUGUGUUGUUAU 2477  
Qy 2474 TCTGATATTGCAAGCTCAAATCTGAGGAAACAAATGAGAGACGGTGAACCCACGNA 2533  
Db 2478 UCCGCAAGGCGGAAACUCAGAAACUCUGCGCAGACUGUCUGGAAACCGGAGAACCGAUGUC 2537  
Qy 2534 CTAATCTCAAGATGTAATCTGTGATGGGTGCTGCTGTGTGGAAGATGACAAAGAGAT 2593  
Db 2538 AGUAGGCAAGGUGUUCUGUGAGCGAGUUCGGGUGUGGAAACCAAGANAU 2597  
Qy 2594 TTTGAAAGATTTGATCTTGATGAGATTTGATCTTGGTTCCTGGAAACAAAGCTGTGT 2653  
Db 2598 CUUUCAGGGUUAUUAUGAAGAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 2657  
Qy 2654 ATGATCAGAGAGGCTAATCTATCTGACTGATAAGAGCCACATGGACATGTGGA 2713  
Db 2658 AUGAUCAGAGAGCGUGCGAAUUCUCCAGGGAUUAUUGUGGCGCACGAGGACACGUAUAA 2717  
Qy 2714 ACGGTAGATTCACTTCTAATGATCCAAACCGCGATCAC-----ACAAGAGGCTT 2764  
Db 2718 ACGUUAUUCUUAUGAUGAAUUAUUGGAAAGACACACGUGUAGUUAUUAUUAUUAUUAUUA 2777  
Qy 2765 TTTATTTGATGAGGGTTGATGTCGACACCGGTTGTGTTAACTTCTGTGTGCTTATCTCT 2824  
Db 2778 UUAUUAUGAAGAGGUGAUGUUGCAUCUGGUGUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 2837  
Qy 2825 GTTGGGACATCGCATACATTTACCGAGATACACAGCAGATTCCTTTCAATTAACAGATT 2884  
Db 2838 UUGUGGAAUUAUGAUGUUAUUGGAGACACACAGCAGAUUCCAUACAUAUAUAGUU 2897  
Qy 2885 CAGAAATTTCCGTATCCCAACATTTTGAGAGAGTGAAGTGGATGAAGTTGAGATGAGG 2944  
Db 2898 UCAGGAUUCUCCUACCCGCCCAUUAUUGCCAAUUGGAUUGAUGAGGUGGAGACACG 2957  
Qy 2945 AGGACACATGAGATGCGCAGGTGATGTAATTTTCTTACATTCGAAGTACGAGGA 3004  
Db 2958 AGAAUUAUCUCUGUGUUGCAGCGCAUGUACACAUUAUCUGAAACAGGAGAUUAGGGC 3017  
Qy 3005 GCGGTGACACCACTTCACTGACCAACGATCGGTCTCATCTGATGATGATGACGCTAG 3064

Db 3018 UUGUUAUGAGCAGUUCUUCGUAUAAAGUUCUUGUUCGAGGAGUUGCGGCGGAGCC 3077  
Qy 3065 GGAGTACTAAACAGTGTTCCTCAAAACCACTAAAGGGGAAATTTGTAATTTTCACTCAGGCT 3124  
Db 3078 GCCGUGAUAUCCGAUCUCAAAACCCUUGCAUGGCAAGAUCCUGACUUAUUAUUAUUAUUA 3137  
Qy 3125 GATAAATTTGATAGGAGAAAGGGCTATAGAAATGTGAACACCGTTTCATGAGATCCAA 3184  
Db 3138 GAUAAAGAGCUCUGCUUUAAGAGGGUUAUCAGAUUGUUCACUGUGCAUGAAGUGCAA 3197  
Qy 3185 GGAGAAACCTTTGAAGATGTCTGCTGTGATGAGATTGACGCAACTCTCACTGCTCTGATT 3244  
Db 3198 GCGGAGACAUACUCUGAUGUUUACUAGUAGGUUAUACCCUACACCAAGUUCUUAUUAU 3257  
Qy 3245 TCCAAGTCTTCCCGCATGTTCTAGTCTCTGCTAGTACACACACAAAGAGCTTCAAAAT 3304  
Db 3258 GCAGGAGACGCCACAUUUGUGCGAUUGUACAGGACACACCUUUGCGCUCAAGUAC 3317  
Qy 3305 TACACCGTAGTGTAGATCTCTTTAGTACAGATAATAGTATGTTGTCTCTTTTAAGCTCC 3364  
Db 3318 UACACUGUUGUUAUGGAUCCUUAUAGUUAUUAUUAUAGAGAUCAUAGAGAAACCUUAGCUG 3377  
Qy 3365 TTCTCTTTAGAAATGTATATGTAAGCAGCTAGTAGTACATGACAAATTCACAGATGATCA 3424  
Db 3378 UACUUGUAGAUUAGUUAUAGUGCAUGCAGAGAACAAUAGCAAUUAUAGAUUAGUUGCU 3437  
Qy 3425 GTGTTCAAAGGTCAATACTCTTTGTGGCAACACCTAAATCAGGAGACTTTTCCAGATCTA 3484  
Db 3438 GUGUUAACAGGUUCCAAUUCUUAUUGUGCAGCGCCAAAGACUGGUGAUUUAUUAUUAUUA 3497  
Qy 3485 CAGTTCTATTAGTATGCTCTCCTGTTATAGTACTATCTTAACTAGTATGATGCT 3544  
Db 3498 CAGUUAUUAUAGUUAUAGUGUCUCCAGGCAACAGCACCAUGAUAUUAUUAUUAUUAUUA 3557  
Qy 3545 GTTACCATGAGTTACGTGATAATAGTCTTAATGTGAAGGATGTGTTCTTGAATTTTCC 3604  
Db 3558 GUUACCAUGAGUUGAUGACAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 3617  
Qy 3605 AAAAGTATTCGATGTCCAAAGAGGTGAAAACATGTCTAGAGCCAGTTTTCGCTACCGG 3664  
Db 3618 AAGUCUGUGCGCCUUAAGAUCAAUCAAACCAUUAUUAUUAUUAUUAUUAUUAUUAUUAU 3677  
Qy 3665 GCGGACCGCAAGGGCTGACGAGTACTCGAATATCTGTTGCAATGATTAAGAAAT 3724  
Db 3678 GCAGAAUUGCCACCGCAGACUGGACUUAUUGAAAUUAUUGUGCGCAUUAUUAUUAUUAUUA 3737  
Qy 3725 TTCACGACACAGACTGACGCGGACGATTGACATTGAGAGCACCGCATCTGTTGTAGTA 3784  
Db 3738 UUAUACGACCCGAGUUGUGGCAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 3797  
Qy 3785 GATAGTTTGTGATGCTATTTTATTAAGAAAGAAATACAAAAAATATTCGTGGA 3844  
Db 3798 GAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAU 3857  
Qy 3845 GTGATGACGAGGATTCAATGATGATGTTGGAAACAGAGAAAGATGATCTATTGGAC 3904  
Db 3858 UUGUUCAGUAGAGAGUCUCUAAUUAUGUGUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 3917  
Qy 3905 GACTTGGCTAACTACAATTTTACAGATCTGCGGCGCATCTGATCAAGACATGATC 3964  
Db 3918 CAGCUGCAGAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 3977  
Qy 3965 AAGGCTCAACCAAAACAGAAATTGACCTTTCAATTCAGATGAATACCTGCTCTGCA 4024  
Db 3978 AAAGCACAACCCCAAGCAAAAUUGGACACUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 4037  
Qy 4025 ACAATTTCTACCATTCGAGCAGATCAACGTTATTTTGGCGGTTCT--CAGAGCTT 4081  
Db 4038 ACGAUUGUGUACCAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 4097  
Qy 4082 ACAAGTTGCTGCTCGAGGCAATTCATTTCTAAGAAAGTTTCTTTTCTTACTAGGAAACT 4141  
Db 4098 ACUAGGCAUUAUCUGGACAGUUGUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 4157







Db 1338 GAUCGUAUCCAGGAGUAUCUAUACGGUGUGACAGCGAGGUCCGAUUGGUAUG 1397  
Qy 1400 GATAAGCAATTCCTCAACCCCTTGTCAATGACTTTCTTCTTGACAGTAAGCTGGCTCG 1459  
Db 1398 GACAAAUUCUUGUAUCAAUCCUUGUCCAGUAGACGUUUUACCUAGUAUUAAGUCCGUG 1457  
Qy 1460 CTTCAAGACGATATAGTAATAGGAAGATTTCGGTCTGGTCTGGTAAGACCACTTCGAAT 1519  
Db 1458 CUAAAGGAUGAUCUAUGUAUAGCAAGUUAUAGUCUGGUAUCCGUAUCCGUGGCCAGCAU 1517  
Qy 1520 ATTTGGGATAGTGGGCAAAATTTTGGAAAGTTTTCGCCACTATCAAGAGAGATTG 1579  
Db 1518 GUGUGGAUGAUAUUCUGCGGCUUUGGGAACGCAUUCUCCUGUAAGAGAGGCGUC 1577  
Qy 1580 GTGACGGAATAATCTGGATGTAAGTGAATGCTCTGAAGATCAAGATCCAGATCTG 1639  
Db 1578 UUGAACAGGAACUAUUCAGAGUGGCGAGCGCAUUAAGAUCAAGGUGGCGUAGUCA 1637  
Qy 1640 TATGTCATCGAAGACAGGTTCTGATGTAATGTAATACCAAGTCTGAGGAGTTACCGAT 1699  
Db 1638 UAUGGACUCCACGACAGAUUAGUGACAGAGUAACAAGGCCUCUGUGGCAUGCCGCG 1697  
Qy 1700 CTAGATATCAAGAGGACTTAGAAGAGCTGAGCAAAATGTACGACGCGTTATCAGAAAT 1759  
Db 1698 CUUGACAUUAGGAAGAGAGUAAGAAACGGGAUGUAUGAUCUAUGCAUUCUUCAGAGU 1757  
Qy 1760 TCTATCTTAAGGGTGTGATAATTTGATATCGGAAGTTTCAAGACATGTGCAAGGCT 1819  
Db 1758 UCGGUGUAAGGAGUCUGACAAAUUCGUAUGUUAUUUCCCAUGUGCCAAUCU 1817  
Qy 1820 TTAGATGTTAGTCTGTGTCGACGACGATGATTCGTTGCGAGTGGCGGAGATAGAAC 1879  
Db 1818 UUGGAAGUGGACCAUAGACGCGCAGCGAAGGUAUAGUCGCGGUAUGAGCAUAGAGAG 1877  
Qy 1880 GGTTTAACTTACTTTTGTATGAAGCCAAACGAGAGAAATGTGGCTAAAGCTCT-----T 1933  
Db 1878 GUCUGACUCUCAUUAUAGCAACGCUACGAGGGAUUGUCCGUAUGUUAUACAGGAU 1937  
Qy 1934 AAAAGACGGGCTGTAGGCGGTGTGTATCTTGAACCGACATCCGAAGAGGTGAAACGTA 1993  
Db 1938 CAAGAGAAGGCUUCAGAGAAGGUGUUGUAGUUAUCCUACAGAGAUGAAGAACCGUCC 1997  
Qy 1994 AATAAATTTCTATTGCTGAGAAAGGAGATTGCTGTGTGTCGAGAAAGTCAAGTTTG 2053  
Db 1998 AUGAAGGUGUUGAUGGCCAGAGGAGAUUCAAUUAAGCUGGUCUUGUGGAGAUCAUCCG 2057  
Qy 2054 ACGAATGTAATTAAGACACAGAGATTGGATCCCTCAACGATTTCCATAAGGCTTGC 2113  
Db 2058 GAGUGCUUAUUAAGAACGAGGAGUAGAGUCUUAUAGAGCGAUUUAUUGCAACG 2117  
Qy 2114 GTGGATAGTGTGATTAACAAGCAATGGCATCGGTTGTCTACCTGGCTCACTCAAGTT 2173  
Db 2118 GCAGAUUCGUAAUUCUAGACAGAGUAGUCUGAUUGUGUACACGGGUCUGAUUAAGUU 2177  
Qy 2174 CAACAATGAAGAACTATGTGACAGTTTGGCAGCTTGTGTCGCGCACTGTATCAAT 2233  
Db 2178 CAGCAAAUAGAAAAUUAUUAUAGCUGGUAUAGCAUACUACUUGCUGCGUGUGCGAAU 2237  
Qy 2234 CTATGCAAGTCAATAAGGATGAGTCCGGTATGATTTCTGATTCAGGGAGAAAGTTGGT 2293  
Db 2238 CUGUCAAGAUCCUCAAAGAUACGUCUUAUUGACCUUUGAAACCCGCUCAAAGUUGGA 2297  
Qy 2294 GTTTGGGATGTCACTTTGAAAAGTGGGCTCTCAAACTGGGGCCAAAGGTCATTTCATGG 2353  
Db 2298 GUCUUGGAUUGCAUUCAGGAUGGUUAUACAACCAACGCGCAAGAGUAGUAGUGG 2357  
Qy 2354 GAGTTGCTGGATTAACAAGGGGAAATGTTTACTGCACTTCTATCTTATGAAGAGAT 2413  
Db 2358 GGUGUUGUAGAAACCCAGCGAGGAAGUAUUAUGUGGCGCUUUGAUAUUAUGAGCAG 2417  
Qy 2414 AGAATGGTCACTGAGAGCGACTGGAGGAGGTTGCTGTATCATCTGTATCAATGGTATAT 2473

Db 2418 GGUGUGUGACAUCCGGAUGAUUUGAGAGAGAGUAGCUGUCAGCUCUGAGUCUGUUGUU 2477  
Qy 2474 TCTGATATTGCAAGCTCAAAATCTGAGGAAAAACAATGAGAGACGGTGAACCCCAAGAA 2533  
Db 2478 UCCGACUGGGCAAAACUCAGAAACUGCGCAGACUGUCUCCGAAACCGAGAAACCGCAUG 2537  
Qy 2534 CTTACTCGAAGATGCTACTTGTGATGGGCTGCTGCTGTTGTGGAAGTACAAAGGAGAT 2593  
Db 2538 AGUAGCGCAAGGUGUUCUUGUGGACCGAGUUCGGGCGUGGGAAAAACCAAGAGAAAU 2597  
Qy 2594 TTTGAAAGATTGATCTTGATGAGGATTGATCTTGGTTCTTGAAAAACAAGCTGCTGCT 2653  
Db 2598 CUUCCAGGGUUAUUAUUGAUGAAGUACUUAUUAUAGUACUUGGGAAGCCGCGGAA 2657  
Qy 2654 ATGATCAGAGAAGAGGGCTTAATCTGATCTGATGATGATGATGATGATGATGATGATG 2713  
Db 2658 AUGAUCAGAGACGUGGGAUUCUCCAGGGAUUAUUGUGGCGACGAGGACAAACGUA 2717  
Qy 2714 ACGGTAGATTCACTTCTTAATGCAT-----CCAAAACCGCATCACAAAGAGGCTT 2764  
Db 2718 ACCGUGAUUCUUAUGAUGAAUUAUUGGGAAGACACACGUCUGACUUAAGAGGU 2777  
Qy 2765 TTTATTGATGAAGGGTTGATGCTGCACACCGGTTGTGTTAACTTCTGCTGCTTATCTCT 2824  
Db 2778 UUCAUUGAUGAAGGUGUAGUGCAUACUGUGUUGUUAUUAUUCUUGGCGGAGU 2837  
Qy 2825 GGTTCGACATCGCATATTTACGAGATACACAGCAGATTCTCTTCAATTAACAGATT 2884  
Db 2838 UUGGCGAAAUUGCAUUGUUAUCCGAGACACACAGCAGAUUCCAUCAUUAUAGAGU 2897  
Qy 2885 CAGAAATTTCCGTTATCCAAACATTTGAGAGCTGCAAGTGGATGAAGTTGAGATGAGG 2944  
Db 2898 UCAGGAUUCGGAUCCCGGCCCAUUAUUGGAAUUGGAGUUGACGAGGUGGAGACGCG 2957  
Qy 2945 AGGACACATGATGCGGATGCGAGTGTGAATTTTCTTCAATCAATCAAGTACGAAGGA 3004  
Db 2958 AGAACUACUCCUUGUCCGAGCGGAGUACAUUAUUGAUCAGAGAGAUUAGAGGCG 3017  
Qy 3005 GCGGTGAACAACCTTCACTGTGACAGATCGGCTCTCATCTGAGATGATAGGCGGTAG 3064  
Db 3018 UUGUCAUGACACUUCUGGUUUAUAAAGUCUGUUGCGAGGAGUUGGCGGAGGCC 3077  
Qy 3065 GGAGTACTTAACAGTGTTCCAACCACTAAAGGGAATTTGTAATTTTCACTCAGGCT 3124  
Db 3078 GCCGUGAUAUCCGACUCUCAAACCCUUGCAUGGCAAGUCCUGACUUUAUCCCAUUG 3137  
Qy 3125 GATAAATTTAGTTAGGAGAGGCTTATAAGAAATGTGAACACCGTTCATGAGATCCAA 3184  
Db 3138 GAUAAAGAGCUCUGCUUUAAGAGGUGUUAUACAGUUAUACACUGGCAUGAAGUGCAA 3197  
Qy 3185 GGAGAAACCTTTGAAGATGTCGCTGCTGATGATGATGATGATGATGATGATGATGATG 3244  
Db 3198 GGCAGACAUAUCUGAUGUUAUACUAGUUAAGGUUAACCCCUACACCAAGUCUCCAUAU 3257  
Qy 3245 TCCAAAGTCTTCCCGCATGTTCTAGTCTGCTGATAGACACACAAAGAGCTTCAAAAT 3304  
Db 3258 GCAGGAGACGCCACAGUUAUUGGUCGCAUUGCAAGGCAACCCUGUUCGCUCAAGUAC 3317  
Qy 3305 TACACCGTAGTGTAGATCTTTAGTACAGATAATTTAGTGTGTTGTTTAAAGTCTCC 3364  
Db 3318 UACACUGUUAUUGGUAUCCUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAU 3377  
Qy 3365 TTCTTTTAGAAATGATATGTTAGAGAGGATGATGATGATGATGATGATGATGATGATG 3424  
Db 3378 UACUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 3437  
Qy 3425 GTGTTTCAAGATCATATCTCTTTGTGGAACACCTTAATCAGGAGACTTTTCCAGATCTA 3484  
Db 3438 GUGUUAAGAGGUCUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAU 3497  
Qy 3485 CAGTTCTTATACGATGATGCTCCCTCGTGAATAGTACTATCTTAAACAAGTATGATGCT 3544  
Db 3498 CAGUUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAU 3557







GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 17, 2005, 06:20:38 ; Search time 341.911 Seconds  
(without alignments)  
13211.223 Million cell updates/sec

Title: US-09-551-494-5  
Perfect score: 6355  
Sequence: 1 gatgttttaatagttttcga.....taaccgcgtagcgccca 6355

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2468	38.8	6395	3	US-09-259-741-1
2	2468	38.8	6395	3	US-09-037-751-1
3	2468	38.8	6395	3	US-09-466-422-1
4	2468	38.8	6395	4	US-09-962-527-1
5	2466.4	38.8	6395	2	US-08-687-559-2
6	2466.4	38.8	6395	4	US-09-401-415-2
7	2455.4	38.6	6439	3	US-09-259-741-2
8	2455.4	38.6	6439	3	US-09-037-751-2
9	2455.4	38.6	6439	3	US-09-466-422-2
10	2455.4	38.6	6439	4	US-09-962-527-2
11	2452.2	38.6	6475	3	US-09-259-741-4
12	2452.2	38.6	6475	3	US-09-037-751-4
13	2452.2	38.6	6475	4	US-09-466-422-4
14	2452.2	38.6	6475	4	US-09-962-527-4
15	2449.8	38.5	6446	3	US-09-259-741-5
16	2449.8	38.5	6446	3	US-09-037-751-5
17	2449.8	38.5	6446	3	US-09-466-422-5
18	2449.8	38.5	6446	4	US-09-962-527-5
19	2431.2	38.3	6425	3	US-09-259-741-3
20	2431.2	38.3	6425	3	US-09-037-751-3
21	2431.2	38.3	6425	3	US-09-466-422-3
22	2431.2	38.3	6425	4	US-09-962-527-3
23	2286.6	36.0	7685	3	US-09-502-710-22
24	2286.6	36.0	7685	3	US-09-502-710-25
25	2286.6	36.0	7685	3	US-09-502-711-22
26	2286.6	36.0	7685	3	US-09-502-711-25
27	2286.6	36.0	7685	4	US-09-565-616A-1

28 2286.6 36.0 7686 3 US-09-502-710-23  
29 2286.6 36.0 7686 3 US-09-502-711-23  
30 2286.6 36.0 7687 3 US-09-502-710-24  
31 2286.6 36.0 7687 3 US-09-502-711-24  
32 2286.6 36.0 7688 3 US-09-502-710-27  
33 2286.6 36.0 7688 3 US-09-502-711-27  
34 2285 36.0 7926 3 US-09-500-554-1  
35 2285 36.0 7926 4 US-09-726-648-1  
36 2285 36.0 7926 4 US-10-119-330-1  
37 2283.4 35.9 7686 3 US-09-502-710-26  
38 2283.4 35.9 7686 4 US-09-502-711-26  
39 2283.4 35.9 7686 4 US-09-565-616A-2  
40 693.2 10.9 1425 1 US-08-488-672-4  
41 690 10.9 1425 4 US-08-708-354-4  
42 302.4 4.8 2173 2 US-08-553-619B-6  
43 270.2 4.3 1825 1 US-08-176-414B-1  
44 270.2 4.3 1825 2 US-08-336-724-1  
45 267.8 4.2 807 4 US-09-565-616A-3

## ALIGNMENTS

RESULT 1  
US-09-259-741-1  
; Sequence 1, Application US/09259741  
; Patent No. 6033895  
; GENERAL INFORMATION:  
; APPLICANT: GARGER, STEPHEN  
; APPLICANT: HOLTZ, R. BARRY  
; APPLICANT: MCCULLOCH, MICHAEL  
; APPLICANT: TURBEN, THOMAS  
; TITLE OF INVENTION: A PROCESS FOR ISOLATING AND  
; TITLE OF INVENTION: PURIFYING VIRUSES SOLUBLE PROTEINS AND PEPTIDES FROM PLANT  
; TITLE OF INVENTION: SOURCES  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Howrey & Simon  
; STREET: 1299 Pennsylvania Avenue N.W.  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA: US/09/259,741  
; APPLICATION NUMBER: US/09/259,741  
; FILING DATE: February 25, 1999  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/037,751  
; FILING DATE: March 10, 1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Halluin, Albert P  
; REGISTRATION NUMBER: 25,277  
; REFERENCE/DOCKET NUMBER: 00801.0140.US01  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-463-8100  
; TELEFAX: 650-463-8400  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 6395 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: unknown  
; MOLECULE TYPE: Genomic RNA  
US-09-259-741-1

Query Match 38.8%; Score 2468; DB 3; Length 6395;

















```

QY 5846 GTGAGATTCTCGATCGGATTTCTATGCTGTATAGATATAATTCGACGCTTGATCCGTTG 5905
Db 5892 GTUAGGUCCGACGAGUGCAUUAAGGUGUACAGGUAACAUGCGGUUAUAGACCCGCUA 5951
QY 5906 ATCAGCGGTTTAAATAGCTTTGATAGTAAATAGATAATAGAGGTTGATAATCAA 5965
Db 5952 GUCACAGACUUGUAGGUGCAUCGACACUAGAAUAGAAUAGAAUAGAAUAGAAUAG 6011
QY 5966 CCGCACCGAATACCTGTAATCTGTAATCGGACCTAGAGGCTAGAGGCTGCTACTGTA 6025
Db 6012 GCGAACCCACGACGCGGAGAGUUAUGUACUCGUGAGUAGAGAGCGACGCGUG 6071
QY 6026 GCTATAAGGCTTCAATCAATTAATTTGCTTAATGAATCTGCTGGAATCGGCATGTC 6085
Db 6072 GCCAUAGGAGCGCAUAAUUAUUAUAGUAAUUAUUAUAGAGGAAACCGGUAUUAU 6131
QY 6086 AATCAAGCAGGCTTTGAGAGCTGTAGTGGACTTGTCTGACACCACTCCGCGCTACTTAG 6145
Db 6132 AUCGAGGACUUCUGAGGACUUCUGUUGUUGUUGGACUUCUGGUCUGCAACU--- 6188
QY 6146 CTATTGTTGTGATTTCTTAAATGCTGCTGAGACTTAAATTCAGGCTGGCTGA 6205
Db 6189 -UGAGGUAGUACAAGUAGUAAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAU 6246
QY 6206 TACCAATATCAGCAGTGTGTTGCTGCTCACTTAAATATAAGATGTCATATCTGATCC 6265
Db 6247 UACGAUAAAGCGUAGUUGUUUUUCCUCCACUUAUUAUUAUUAUUAUUAUUAUUAUUA 6305
QY 6266 AACAGTTAAACCATGATGTGTGTACTGTGTGCTGTAATGCGTAAACATCGGAGGTTTCG 6325
Db 6306 GCGGUGUAAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 6365
QY 6326 AATCTCCCTTAAACCGCGGTAGCGGCCA 6355
Db 6366 AAUCCCGCGUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 6395

```

# RESULT 3

```

US-09-466-422-1
; Sequence 1, Application US/09466422
; Patent No. 6303779
; GENERAL INFORMATION:
; APPLICANT: GARGER, STEPHEN
; HOLTZ, R. BARRY
; MCELLOCH, MICHAEL
; TURPEN, THOMAS
; TITLE OF INVENTION: A PROCESS FOR ISOLATING AND
; PURIFYING VIRUSES SOLUBLE PROTEINS AND PEPTIDES
; FROM PLANT SOURCES
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howrey & Simon
; STREET: 1299 Pennsylvania Avenue N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/466,422
; FILING DATE: 17-Dec-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/037,751
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Halluin, Albert P
; REGISTRATION NUMBER: 25,277

```

REFERENCE/DOCKET NUMBER: 00801.0140.999

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-463-8109

TELEFAX: 650-463-8400

TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 6395 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: unknown

MOLECULE TYPE: Genomic RNA

SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-466-422-1

Query Match 38.8%; Score 2468; DB 3; Length 6395;

Best Local Similarity 44.6%; Pred. No. 0;

Matches 2851; Conservative 1173; Mismatches 2300; Indels 66; Gaps 11;

```

QY 20 ACAACAACAATTAACAAACAAACAAACAAATATTACAAACAACAACAACAACAACAATGGCAC 79
Db 18 ACCAACACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 77
QY 80 ATACAATCTATAATTAGCAACGCCCTTCTTTGAAAGCTGAGTGTGTAATAAACTCTCGTT 139
Db 78 ACACAGACAGCUACCAACAUCAUCAGCUUUGCUGACACUCUGCCGAGGAAACAACUCCUGUC 137
QY 140 AATGACCTTGTCAAGAGGCGCATGTACGATACGCGCTGGAAGAAATTTAAGCCCGCGGAC 199
Db 138 AUAUACUAGCAAGCGUUCUUAACACACAGCGGUAGAGUUAUUAAGCUCUGAC 197
QY 200 CGTAGACAAAGGTCATTTTCCAAAACCTATTAGCGAAGAGCAACGCTTCTAGTCTCC 259
Db 198 GCGAGGCCAAGGUGAACUUUCAAAGAAUUAAGCGAGGAGAGCGCUUUAUUGCUAC 257
QY 260 AACGCTTACCGGAGTTCAGATTTCTTTATATCTCAAAATGCGGTACACAGTTTG 319
Db 258 CGGCGUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 317
QY 320 GCTGAGGTTTGAGACATTTAGAAATGGAATATCTGATGCTCAAGTTTCCCTATGGATCG 379
Db 318 GCAGUGGUAUUGCGUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAU 377
QY 380 CCGACATATGATAGTGGGAACTTTGCGACACATTTGTTCAAAGGAGGATTCGTG 439
Db 378 UUGACUUAUGACUAGCGGGAUUUUGCAUCGCUUUAUUAUUAUUAUUAUUAUUAUUAUUAU 437
QY 440 CATGCTGTATGCGCCATCTGACATACGAGATATATAGCGACGAGGACAAAGGAC 499
Db 438 CACUGCUGACGCGCCACCCUGGACUUCGAGCAUCAUUGCGGACGAGGAGGAGGAGGAGG 497
QY 500 TCAATTGAGATGTATTTGTCAGATTTGTCTCGTTCTAAAGGTAATTTCTGAGTTTCAA 559
Db 498 AGUAUUGAACUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAU 557
QY 560 AGGAGGCTTTTAACAGGTATGAGAGCTCCCAACGAGTCTGCTCTTAAACATTTT 619
Db 558 AAGGAAGCAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAU 617
QY 620 CAGGATTTGCGAATAATCATCCGCGAGAGAAATAGTGTAGAAATAGCGTTGCTGTCAC 679
Db 618 CAGACAUGCGACAUCAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAU 677
QY 680 AGTTTGTATGATATTTCTGTGATGAGTTTGAGTGTGTTTAAATATCTAAGATATACAT 739
Db 678 AGCAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 737
QY 740 GTATGTTATGAGCTTCCATTTTGCAGAAACATTTACTAGACACGAGGAGTTAG 799
Db 738 ACGUGCUAUGCGCGUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 797
QY 800 CTTAATGAATAGCGCAACTTTCAAAGAGAGAGGTGATGTTCTTTTCTTTCTTCTGCT 859

```



Db 798 UUGAGCGAAUUCACCGCGUGUUUUUCGCGGAGUGAGACAAAGUUUGACCUUUUUUUUGCA 857  
Qy 860 GATGAAAGTACTTTAAATATTAGTCATAAAATACAAAATATCTTGCAATTATGTAGTTAAA 919  
Db 858 UCAGAGAGUACUUAUUUUGUACUAGUUUAUCUAAUUCUUAAGUUGUGUGCAAA 917  
Qy 920 TCTTACTTTCTGCTTCTAGTAGTAATAGTTTACTTTTAAAGAAATTTTATAGTCACTAGGGTT 979  
Db 918 ACUUAUUCUCCGCGCUCUUAUAGAGAGUUUAUCAGAAAGAGUUUUUAGUACACAGAGUU 977  
Qy 980 AATACTTGGTTTCTTAAATTTACCAAGTAGATACCTATATCTTGTAACAAGAGTTTAGA 1039  
Db 978 AAUACCUUGUUUUUGUAGUUUUUUGAAGUAGUACUUUUUUUUUGAACAAGGUGUGGCC 1037  
Qy 1040 CAAGTAGGGTGTAGTAGTATAGTCTTCTATGAGCGATGGAAGACGCTTTTGTCTTACAAG 1099  
Db 1038 CAUAAAAGUGUAGUAGUAGACAGUUUAUCUGCAUUGGAAGACGCAUGGCAUUACAAA 1097  
Qy 1100 AAAACCTTGGCCATGTTCAACACTGAAGAGCAATCTTTAGAGACACGGCTTCGGTTAAC 1159  
Db 1098 AAGACUCUUGCAUUGUGCAACAGCGAGAGAAUCCUCCUUGAGGAUUCUAUCAAGUCAAU 1157  
Qy 1160 TTTTGGTTCCCTAAGATGAAGGACATGCTGATAGTACCGCTGTTTGGGGTCTTATTACC 1219  
Db 1158 UACUGUUUUCCAAAAGAGGGAUUGUGUACUGUACCAUUUUCGACAUUUUUUGGAG 1217  
Qy 1220 AGCAAAAAGATGACAAAGGAGTGAAGTCAATCTGTTAATCGTGACTTCTGTTTACACAGTGT 1279  
Db 1218 ACUAGUAGAGGACGCGCAAGAGUACUUGUCCAAAGGAUUUGUUGUUUAACAGUCUU 1277  
Qy 1280 AATCATATCAGAACATATCAAGCCAAAGCGTTAACTTACCAAGAGTATATTCTTTCGTG 1339  
Db 1278 AACCACAUUCGAAUACACAGCGCAAGACUCUUAUCAACGCAAAUUGUUUUGCCUUGUC 1337  
Qy 1340 GAGTCTATAGATCCCGTGATATCAATCAATGCTGTACTGCTAGCTCTGATGCGATGTA 1399  
Db 1338 GAUCGUAUUCGAGGAGUAAUUAUUAACGUGUGACAGCGAGGUGCCGAAGGGAUGUG 1397  
Qy 1400 GATAAGCAATTTCTTCAACCCCTGCTCAATGACTTTCTTCTTGACAGACTAAGCTGGCTGCG 1459  
Db 1398 GACAAUUCUUGUUAUCAAUCCUUGUCCUAGAGUUUAUACUGCAUUAUAGUCCUGCGUU 1457  
Qy 1460 CTTCAAGACGATATAGTAATGGGAAAGTTTCGGTGTGTTGGATAGACACATCTTGAATTT 1519  
Db 1458 CUAAAGGAUGACUUAUGUAGCAAGUUUAGUUCUGUUCGUAAGAAACGUGUGCCAGCAU 1517  
Qy 1520 ATTTGGATGAGTGGGCAAAATTTTGGAAACGTTTCCCACTATCAAGAGAGATTG 1579  
Db 1518 GUGUGGAUGAGAUUUCGUGCGGUUUGGGAACGCAUUUCCUCCGUGAAAGAGACUC 1577  
Qy 1580 GTGACGAGGAAATTTCTGATGTAAAGTGAAGTCTCTGAAGATCAAGATCCAGATCTG 1639  
Db 1578 UUGAACAGGAACUUAUCAGAGUGCGGCGACGCAUUAUAGAGUACAGGUGCCUGAUCUA 1637  
Qy 1640 TATGTACATGGAAGACAGGTTCTGTAGTCAATACCAAGTCTGAGAGTTTACCGCAT 1699  
Db 1638 UAUGUGACUUCACGACAGAUUAGUAGACUAGUACAGGCGCUCUGUGGCAUGCCUGCG 1697  
Qy 1700 CTAGATATCAAGAGGACTTGAAGAGACTGAGCAAAAGTACGACGCGTTATCAGATTATTA 1759  
Db 1698 CUUGACAUUAGGAAGAGUAGAGAAACGGAAGUGAUGAUCUUAUGCAUUCUUCAGAUUA 1757  
Qy 1760 TCTATCTTAAAGGTGCTGATAATTTCCGATATCGCAAGTTTCAAGACATGTGCAAGGCT 1819  
Db 1758 UCGGUGUUAAGGAGUCUGACAAUUCGAGUUGAUGUUUUUCCAGAGUUGCCNAUCU 1817  
Qy 1820 TTAGATGTTAGTCTGTAGTGTGGACGACGAGTAATCGTTGAGTGGCGGAGAAATAGAAC 1879  
Db 1818 UUGAAGUUGUACCAUAGACGCGCAAGGUAUAGUAGUGCGGUCUAGAGCAUAGAGAGC 1877  
Qy 1880 GGTTTAACTCTTACTTTTATAGCCACCGAGAGATGTGGCTTAAGGCTCT-----T 1933  
Db 1878 GGUTGACUCUCAUUUUGAACGACCUACUGAGCGGAUGUUGCGUAGCUUAUAGGAG 1937

Qy 1934 AAAAGCACGCGCTGTGAGCGCGTGTGTATGTCTTGAACCGACATCCGAAAGAGGTGAACGTA 1993  
Db 1938 CAAGAGAAGGCUUCAGAGAAGGUGCAUUGGUAGUUUACCUCAAGAGAAGUUGUAGAAGACCGUCC 1997  
Qy 1994 AATAAATTTTCTATTGCTGAGAAAGGAGATTGCTGTGTGTGCAGAAAGTCATGGTTTG 2053  
Db 1998 AUGAAGGUGUGGAGCGGCGAGAGAGAUACAUAUUGUGUGUCUUGUGGAGAUCAUCCG 2057  
Qy 2054 ACGAATCTAACTTATAGAGCACGAGAGTTGAGAGTCCCTCAACGATTTTCCATAAGGCTTGC 2113  
Db 2058 GAGUGUCUUAUUCUAAAGACGAGGAGAUAGAGUUCUUAAGAGAGUUAUUAUUGGCGAGC 2117  
Qy 2114 GTGATAGTGTGATTAACAAGCAAAATGGCATCGGTGTCTACATCTGGCTCACTCAAAGTT 2173  
Db 2118 GCAGAUUCGUUAUUCGUAAAGCAGUAGCGUAGCGUUAUCAAACGCGCAAGAGCAUUGAAU 2177  
Qy 2174 CAACAATGAAGACTATGTGGACAGTTTGGCAGTTTGGCAGCTTCTGTGTCTGCCACCTGTATCAAT 2233  
Db 2178 CAGAAAUGAAAAACUUUUCGUAUGCCUGUAGUACUUAUCUGUGCGUGUGUGUAGAU 2237  
Qy 2234 CTATGCAAGTCACCTAAAGGATGAAGTCGGGTATGATTTCTGATTTCCAGGGAGAAAGTTGGT 2293  
Db 2238 CUCGUCAAAGUCCUCAAAGAUACAGCUGCUUAUUGACCUUGAACCUCUCAAAGAUUUGGA 2297  
Qy 2294 GTTTGGGATGTCACTTTTGA AAAAGTGGCTCTCTCAAACTCGCGCCAAAGGTCATTCATGG 2353  
Db 2298 GUCUUGAUGUGCAUCUAGAGAGUGGUAAUCAAACCAACGCGCAAGAGCAUUGCAUGG 2357  
Qy 2354 GGAGTGTCTCGGATTAACAAGGGGAAAATGTTTACTGCATCTTCTATCTTATGAAGGAGAT 2413  
Db 2358 GGUGUUGUAGAAACCCAGCGGAGAGUUAUCAUGUGGCGCUUUUGGAAUAUGAUGAGAG 2417  
Qy 2414 AGAATGTGATCTGAGACGCACTGAGAGAGGTGCTGTATCATCTGTATCAATGATATAT 2473  
Db 2418 GGUGUGUGACUCCGAGUAGUUGGAGAGAGUAGUACUUGUAGCUCUGAGUCUGUUGUUAU 2477  
Qy 2474 TCTGATATTGCAAGCTCCAAAATCTGAGGAAAAACAATGAGAGACGGTGAACCCCAAGAA 2533  
Db 2478 UCCGACAUUGCGAAACUCAGAAACUCUGCGCAGACUGUCUUGGAAACCGAGAAACCGCAUGC 2537  
Qy 2534 CCTACTCAAGATGTTGATCTTGTGATGGGTGCTGTGTTGGAAGTACAAGAGGAGAT 2593  
Db 2538 AGUAGCCCAAGGUGUUCUUGUGGACGAGUUCUGGCGUGUGGAAACCAAGAGAAU 2597  
Qy 2594 TTTGAAAGATTTGATCTTGTGATGAGATTTGATCTTGTGTTCTGGAACCAAGCTGCTCT 2653  
Db 2598 CUUCCAGGUAUUAUUGUAGAGAUUAUUAUUGUACCUUGGAGAGCAAGCCGCGAA 2657  
Qy 2654 ATGATCAGAGAAGGGCTAATTTCACTGGAAGTGAAGAGCCACAATGGAACAATGTGAGA 2713  
Db 2658 AUGAUCAGAAAGCUGGCGAAUUCUCCAGGGAUUAUUGUGGCCACGAAAGGACAACGUAUAA 2717  
Qy 2714 ACGTATGATTCATCTTCTAATGCAAT-----CCAAAACCGCGATCACAAAGGGCTT 2764  
Db 2718 ACCGUAGUUCUUAUGAUGAAUUUUGGGAAGACACACGUGUGUUAUCAAAGGUAU 2777  
Qy 2765 TTTATGATGAAGGGTTGATGCTGCACACCGGTTGTGTAACTTCTCGGTGCTTATCTCT 2824  
Db 2778 UUCAUUGAAGAGGUGUAGUUGCAUCUUGUUGUUAUUAUUCUUGUGCGGAGUUA 2837  
Qy 2825 GGTTCGACATCGCATACATTTACGGAGATACACAGCAGATTTCTTTTCAATTAACAGATT 2884  
Db 2838 UUGUGCGAAAUUGCAUUGUUUAUCGAGACACACAGCAGAUUCCAUACAUCAAUAGAGU 2897  
Qy 2885 CAGATTTCCGTTATCCCAACATTTTGAAGAGTGCAGAGTGGATGAGTTGATGAGG 2944  
Db 2898 UCAGAUUCCGUAACCCCGCCCAUUAUUGCCAAUUGGAAUUGGAGUUGGAGGAGGACACGC 2957  
Qy 2945 AGGACCACTGAGATGCCAGGTGATGTAATTTTTTCTCAATCAATCAAGATCAAGAGGA 3004  
Db 2958 AGAAUACUCUCCGUUGCCAGCGGAGUACAUUAUUGAACAUAUUGACAGAGAUUAGGAGC 3017



Db 5175 UCUGUGGACAAAAGGAUGGAAAGCGCAGCGCCACUCUCUGGAUCUUAUACACAGC 5234  
Qy 5222 TGCTTGCAAAAGAAATTTCTTTTAAAGCTAATCCCTAATTAATTAATCAATAACATCCGAGGA 5281  
Db 5235 AGCUGCAAAAGAAAGAUUACUGUUAUACGAGGUGUCCCAUUAUGCUUAUACCAACCCAGGA 5294  
Qy 5282 TGCTGAGNAGCACCGCTGGCAAGTGTGTAGTGAATCAAGAGAGTGGCTATGGAAGAGG 5341  
Db 5295 CGCAUGAAGAAACGUCUGCGCAAGUUAUUAUUAUUAAGAAUUGGAGUAGUUGCAGCGGG 5354  
Qy 5342 ATACTGTCCTTTATCTTTGAGAGTTCGTTTCAATTTGTGTAGTACATATAAAATAATGTAAAG 5401  
Db 5355 UUUCUGCCGCUUUCUGUGAGUUGUGUGCGUGUGUUAUUGUUAUUAAGAAUUAUUAUAA 5414  
Qy 5402 AAAAGTTTGAAGGAAACGTATTTTGAAGTGTGACAGACGGCTCGGCAATTTGAACCTACTGA 5461  
Db 5415 AUUAGGUUUGAGAGAGAGAAUUAUACAAACUGAGAGACGGAGGCCCAUGGAACUUAACAGA 5474  
Qy 5462 AAAGTTTGTGAGGAGTTCGTTGATGAGTAGTACCAATGGCTGTGAACCTCGAAAGGTTTC- 5520  
Db 5475 AGAAGUCUGUAGAGUUAUGGAAGAUUGUCCUUAUGUGCAUGAGGCUUGCANAGUUGC 5534  
Qy 5521 -----CGGAAACAAAAGAAATGGTAGGTAATAATGTTTAATAAT----- 5561  
Db 5535 AUCUGRAACCGGAAAGAGUGAUGUCCGCAAGGGGAAAUAGUAGUAGUAGUGGUC 5594  
Qy 5562 -----AAGAAATAATAACAGTGTGTAAGAGGGTTTTTAAATTTGAGGAATTTGAGGA 5614  
Db 5595 AGUGCGGAACAAGAAUUAUAGAAUUGUUAAGAUUUGGAGGAUAGUUAUUAAGAAAGAA 5654  
Qy 5615 TAATGTAGTATGACAGAGT-----CTATCGCTCATCGAGTAGGCTTTTATCAAT 5665  
Db 5655 UAAUUAUACUGAUGAUGAUGCGAGGCUACUGUGCGCAUUGCGGAUUGUUAUUAUUA- 5711  
Qy 5666 ATGCTTATACAACTCAACTCTCGAGCAATTTGTTTACTTATCTTCGCTTACGCGAGAT 5725  
Db 5712 AUGUUAUACAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 5771  
Qy 5726 CCTGTGACGTGATCAATCTGTGTGTAACAAATGATTTGGTAAACCGATTTTCAACGCAACAA 5785  
Db 5772 CCAUAGAGUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 5831  
Qy 5786 GCTAGGACACAGTCCACACAGCAATTTGCGGATGCGTGGAAACCTGTGCTGATGATGACA 5845  
Db 5832 GCUGAAACUGUGUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 5891  
Qy 5846 GTGAGATTTCTGATCGATTCGTTCTATGTGTATAGATATAATTCGACCTTTGATCCGTTG 5905  
Db 5892 GUUAGGUUCCUGACAGUGACUUAUAGGUGUACAGGUACAUUGCGUUAUUAAGACCCGCUA 5951  
Qy 5906 ATCAGCGGTTTAAATAGCTTTGATAGTAAGAAATAGAAATAGAGGTTTGAATATCAA 5965  
Db 5952 GUCACAGACUUGUAGUGCAUUGACACUAGAAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 6011  
Qy 5966 CCGCAGCGATATCTAGTGAATCGTTAACCGGACTAGAGGTTAGACGATGCTACTGTA 6025  
Db 6012 GCGAACCCTCAGACGUGUAGUAGUAGUAGUAGUAGUAGUAGUAGUAGUAGUAGUAGUAGU 6071  
Qy 6026 GCTATAAGGCTTCAATCAATAATTTGGCTAATGAATGCTGTTGGAATCGGCATGTTTC 6085  
Db 6072 GCUUAGGAGCGGCAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 6131  
Qy 6086 AATCAAGCAGGCTTTGAGAGCTGCTAGTGGACTTGTCTGAGCACCAACTCCGGCTACTTGA 6145  
Db 6132 AAUCGGAGCUUUCGAGAGCUUUCUGUGUUGUUGUUGUUGUUGUUGUUGUUGUUGUUGUUGU 6188  
Qy 6146 CTATTGTTGAGATTTCTTAAATAAGTCTGTAAGACTTAAATTTACGGTGGCTGA 6205  
Db 6189 -UGAGGUGAUGCAUGAUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 6246  
Qy 6206 TACCAAAATCAGAGTGGTGTGCTGCACCTTAAATATAAGATTTGTCATATCTGATCC 6265  
Db 6247 UACGAUAACGCAUAGUGUGUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 6305

Qy 6366 AACAGTTAAACCATGTGATGTGTATCTGTGTATGCGGTAAACATCGGAGAGGTTTCG 6325  
Db 6306 GCGGUCAAAUUGUAUUGGUUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 6365  
Qy 6326 AATCCTCCCTAAACCGCGGTAGCGGCCA 6355  
Db 6366 AUCCCCCGUUUACCCCGUAGGGGCCA 6395

## RESULT 4

US-09-527-1  
; Sequence 1, Application US/09962527  
; Patent No. 6740740  
; GENERAL INFORMATION:  
; APPLICANT: GARGER, STEPHEN  
; HOLTZ, R. BARRY  
; MCCULLOCH, MICHAEL  
; TUREN, THOMAS  
; TITLE OF INVENTION: A PROCESS FOR ISOLATING AND  
; PURIFYING VIRUSES SOLUBLE PROTEINS AND PEPTIDES  
; FROM PLANT SOURCES

; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Howrey & Simon  
; STREET: 1299 Pennsylvania Avenue N.W.  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/962,527  
; FILING DATE: 24-Sep-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/037,751  
; FILING DATE: 10-march-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Halluin, Albert P  
; REGISTRATION NUMBER: 25,277  
; REFERENCE/DOCKET NUMBER: 00801.0140.999  
; TELEPHONE: 650-463-8109  
; TELEFAX: 650-463-8400  
; TELEX: <Unknown>  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 6395 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: unknown  
; MOLECULE TYPE: Genomic RNA  
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-527-1

Query Match 38.8%; Score 2468; DB 4; Length 6395;  
Best Local Similarity 44.6%; Pred. No. 0;  
Matches 2851; Conservative 1173; Mismatches 2300; Indels 66; Gaps 11;

Qy 20 ACAACAACAATTAAAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 79  
Db 18 ACCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 77  
Qy 80 ATACAATCTATAATTAGCAACGCCCTTTCTTTGAAAGCGTGTAGTGGTAAACACCTCTCGTT 139  
Db 78 ACACAGACAGCUACCAUCAUCAGCUUUGUGGACACUCUGCCGAGGAAACAACUCCUGGUC 137  
Qy 140 AATGACCTTGCAGAAGAGCGCATGTACGATACGCCCGTGGGAAGAAATTTAACGCCCGCAGC 199





[illegible]

Db	5595	AGUGCCGACAAGAA	CUUUAUGAAAUAGUUAAGNAUUUGGAGGAAUGAGUUUUUAAAAGRA	5655
Qy	5615	TAATGTAAGTGATGACGAGT	-----CTATCGCGTGCATCGAGTACGTTTAAATCAAT	5665
Db	5655	UAAUUAAUCGAUGAUUCGGAGGACUACUGUCGCCGAUUCGUAUUCGUUUUA	---AAU	5711
Qy	5666	ATGCTTTATCAATCAACTCTCGAGCGCAATTTGTTTATCTTATCTTCCTCGCTTACGCAGAT		5725
Db	5712	AUGUCUUAAGUAUCAUACUCCAUUCGACUGUCUGUCUUGUAUCAGCGUGGGCGCAG		5771
Qy	5726	CCGTGTCAGCTGATCAATCTGTCTGACAAATGCAATTCGGGTAAACAGTTTCAAAACGCAACAA		5785
Db	5772	CCAAUAGAGUUAUUAUUAUGUAUUAUGCCUUAAGAAUACAGUUUCAAACACACAA		5831
Qy	5786	GCTAGGACAAACAGTCCAACAGCAATTTGCGGATGCTCGTGGAAACCTGTGCTAGTAGACA		5845
Db	5832	GCUCGAACUGUCUUCAAAGACAAUUCAGUGAGGUGUGGAAACCUUCAACCAAGUAUACU		5891
Qy	5846	GTGAGATTTCTGTCATCGGATTTCTATGTGTATAGATATAATTCGACGCTTGATCCGTTG		5905
Db	5892	GUUAGGUUCCUGACAGUGACUUUAAGGUGUACAGGUACAAGUCGGUUAUJAGACCCGCUA		5951
Qy	5906	ATCACGCGCTTTAAATAGCTTTGATCTAGAAATAGATAATATAGAGTTGATATAATCAA		5965
Db	5952	GUACACGACUGUUGGUGCAUUCGACACUAGAAAUAGAAUUAAGUAGAAAUACAG		6011
Qy	5966	CCCECACCGCAATACTACTGAATTCGTTAAACGCACTCAGAGGGTAGACGATGCTACTGTA		6025
Db	6012	GCGAACCCACGACGUCGCGAGAGCGUUAUGAUGCUACUCGUAGAGUAGACGACGCAACGGUG		6071
Qy	6026	GCTATAAGGCGTTCAATCAATATTTGGCTTAATGAACTGGTTTCGTGGAACCTGCGATGTTT		6085
Db	6072	GCCAAUAGGAGCGCGAUAUAUAUUAUUAUGAUGAAUUGAUCAGAGAAACCGGAUCUUU		6131
Qy	6086	AATCAAGCAGGCTTTGAGACTGCTAGTGGACTTGTCTGTGACCAACAACCTCCGGTACTTAG		6145
Db	6132	AUUGGAGCUUUCUUGAGAGCUCUUCUGGUUUGUUGAGCCUUCUGUCCUGCAACU---		6188
Qy	6146	CTATTGTTGTGAGATTTCTTAAATAAAGTCGCTGAAGACTTTAAATTCAGGGTGCTGA		6205
Db	6189	-UGAGGUAGUCAAGUACAUAUAAUAAUACGGAUUGUGCCGUUAUACACA-CGUGGUGCG		6246
Qy	6206	TACCAAAATCAGCAGTGTTGTTTCGTCACCTTAAATATACAGATTGTCATATCTGATCC		6265
Db	6247	UACGAUACGUAUGUGUUUUUCCUCCACUUAUACGAAUGAGGGUUGU-GUCUUGGAGUCG		6305
Qy	6266	AAACAGTTAAACCATGTGATGGTTATCTACTGTGTGTGGCGTAAACATCGGAGAGTTCG		6325
Db	6306	CGCGUCAUAUGUAUUGGUUCAUAUACAUCCGCAGGCACGUAUAAAGCGAGGGGUUCG		6365
Qy	6326	AATCTCTCCCTAACCCGCGGTAGCGGCCA		6355
Db	6366	AUCCCCCGTUAACCCCCCGUAGGGGCCA		6395

RESULT 5  
US-08-687-559-2  
; Sequence 2, Application US/08687559  
; Patent No. 5955647  
; GENERAL INFORMATION:  
; APPLICANT: Fitchen, John H.  
; APPLICANT: Beachy, Roger N.  
; TITLE OF INVENTION: METHOD FOR USING TOBACCO MOSAIC VIRUS TO  
; TITLE OF INVENTION: OVERPRODUCE PEPTIDES AND PROTEINS  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson, P.C.  
; STREET: 4225 Executive Square, Suite 1400  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: US  
; ZIP: 92037  
; COMPUTER READABLE FORM:







Db 1698 CTTGACATTAGAGAGATCGAAGAACGGAAGTGAATGATCAATGCACATTTTCAGAGTTA 1757  
Qy 1760 TCTATCCTTAAGGGTCTGATAAATTTTCATATCGGAAGTTCAAGACATGTGCAAGCT 1819  
Db 1758 TCGGTGTTAAGGAGTCTGACAAATTCGATGTGATGTTTTTCCGAGATGCGCAATCT 1817  
Qy 1820 TTAGATGTTAGTCTGATGTCGACGACGAGTAAATCGTTGAGTGGCGGCGAATAGAACG 1879  
Db 1818 TTGAAGTTGACCCATGACGCGACGAGGTTATAGTCGGGTCTATGAGCAATGAGAGC 1877  
Qy 1880 GGTTTAACTCTTACTTTTATAAGCCAAACGAGGAGAAATGTTGCTTAAGGCTCT-----T 1933  
Db 1878 GGTCTGACTCTACATTTTGAACGACTACTGAGCGAATGTTGCGTAGCTTTTACAGGAT 1937  
Qy 1934 AAAAGCAGCGGTCTGAGCGGTGTATGTCTTGAACCGCATCCGAAGAGGTGAACGTA 1993  
Db 1938 CAAGAGAAGGCTTCAGAAAGGTCTTTGGTAGTTACCTCAAGAGAAATTTGAAGAACCGTCC 1997  
Qy 1994 AATAAATTTTCTATTGCTGAGAAAGGAGATTTGCCCTGTGTGCGAAGATCATGGTTG 2053  
Db 1998 ATGAAGGTTCAATGCCAGAGGAGATTCAATTAGCTGTCTTGTGGAGATCATCG 2057  
Qy 2054 ACGAATGCTAACTTAGACACACAGGAGTTGGAGTCCCTCAACGATTTCCATAAGCTTGC 2113  
Db 2058 GAGTCGTCTTATCTAAGAACGAGGAGATAGAGTCTTTTAGAGCAGTTCATATGGCAACG 2117  
Qy 2114 GTGGATAGTGATTAACAAGCAATGATCGTGTCTACACCTGGCTCACTCAAGTT 2173  
Db 2118 CGAGATTCGTTAAATTCGTAAGCAGATGAGTCTGATTGTGTACACGGGTCCGATTAAGTT 2177  
Qy 2174 CAACAAATGAAGAACTATGTGACAGTTTGGCAGCTTCGTTGTCGCGCACTGTATCAAT 2233  
Db 2178 CAGCAATGAAATCTTTATCGATAGCTGTGATGATCACTATCTGCTCGGTGTGGAAT 2237  
Qy 2234 CTATCGAAGTCACTAAGAGTGAAGTCTGGGTATGATTTCTGATTCAGGGAGAAAGTTGGT 2293  
Db 2238 CTCGTCAAGATCCTCAAGATACAGCTGCTATTGACCTTTGAACCCCGTCAAAAGTTTGA 2297  
Qy 2294 GTTTCGGATGTCATTTGAAAGTGGCTCCTCAACCTCGCGGCAAAAGTCAATTCATGG 2353  
Db 2298 GTCTGGATGTTGCATCTAGGAAGTGGTTAATCAACCAACGCGCAAGTCACTCATCG 2357  
Qy 2354 GGAGTGTCTCGATTTACAAGGGGAAATGTTTACTGCACTTTCTATCTTATGAAGAGAT 2413  
Db 2358 GGTGTTGTGAACCCACGAGGAGTATCATGTGGCGCTTTTGAATATGATGAGCAG 2417  
Qy 2414 AGAATGTTGACTGAGAGCAGTGGAGGAGGTGGCTGTATCATCTGATACATGTTAT 2473  
Db 2418 GGTGTGTTGACATGCGATGATTGGAGAGAGTAGCTGTGAGCTCTGTTGTTAT 2477  
Qy 2474 TCTGATATTGCAAGCTCCAAATCTGAGGAAACAAATGAGAGACGTTGAACCCACGAA 2533  
Db 2478 TCCGATGCGGAAACTCAGAACTCTGCCAGACTGCTCGAAGCGGAAACCGATGTC 2537  
Qy 2534 CTTACTGCAAGATGTTACTTTGGATGGGGTCCCTGGTTGTGGAAGTCAAAAGAGAT 2593  
Db 2538 AGTAGCGCAAGGTTCTTCTTGGACGAGTTCGCGGTGTGGGAAACCAAGAAAT 2597  
Qy 2594 TTTGAAAGATTTGATCTTATGAGGATTTGATCTTGGTTCTTGGTGAAGAACAGCTGCT 2653  
Db 2598 CTTTCCAGGGTTAAATTTGATGAAGATCTAAATTTAGTACCTGGGAAGCAAGCCGGA 2657  
Qy 2654 ATGATCAGAAGAGGCTTAATTCATCTGACTGATAAGAGCCCAATGGAATGAGTGA 2713  
Db 2658 ATGATCAGAAGAGCTCGAATCTCTCAGGATTTATTTGTGCCACGGAAGCAACGTTAA 2717  
Qy 2714 ACGGTAGTTCACTCTTAATGAT-----CCAAACCGGATCACAAGAGGCTT 2764  
Db 2718 ACCGTTGATCTTTCATGATGAATTTTGGGAAAGCACACGCTGTCAAGAGGTTA 2777  
Qy 2765 TTTATTTGATGAGGTTGATGTCACACGGTTGTGTTAACTTCTCGTGTCTTATCTCT 2824

Db 2778 TTCATTGATGAAGGGTTGATGTTGCATACCTAGTGTGTGTTAAATTTTCTTGTGGCATGTCA 2837  
Qy 2825 GGTTCGACATCGCATATTTACGGAGATACACAGCAGAGATTCCTTTCAITAAACAGAGTT 2884  
Db 2838 TTGTGCGAATTTGCATATGTTTACGGAGACACACAGCAGATTCATACATCAATAGAGTT 2897  
Qy 2885 CAGAAATTTCCGTATCCCAAACTTTTGAAGCTGCAAGTGAAGTGAAGTGAAGTGAAG 2944  
Db 2898 TCAGGATTTCCGTACCCCGCCCATTTTGCCTAAATTTGAAGTTGACGAGGTGGAGACGC 2957  
Qy 2945 AGNACCACTGAGATCCCGAGGTGATGTAATTTTCTTACAAATCGAAGTACGAAGGA 3004  
Db 2958 AGAATCTCTCTCGGTTGTCCAGCGGATGTCAATTTCTGAAACAGAGATATGAGGCG 3017  
Qy 3005 GCGGTGACAACTCTCAACTGTACAACTGTAACAGTCTCATCTGAGATGATAGCGGTAAG 3064  
Db 3018 TTTGTCTAGGACTTCTTCGGTTAAAGTCTGTTTCGAGGAGATGGTCCGCGGAGCC 3077  
Qy 3065 GGAGTACTAAACAGTGTTCCAAACCACTAAAGGGAATTTGTAACCTTTCACTCAGGCT 3124  
Db 3078 GCGGTGATCAATCCGATCTCAAAACCCCTTGATGGCAAGATCCCTGACTTTTACCCAAATCG 3137  
Qy 3125 GATAAATTTGAGTTAGAGGAGAGGCTATGAAGATGTGAACACCGTTTCATGAGATCCAA 3184  
Db 3138 GATAAGAGAGCTCTGCTTTCAAGAGGTATTCAGATGTTCACTGTGCAATGAAGTCAA 3197  
Qy 3185 GGAGAAACCTTTGAAGATGTTGCTGCTGATGAGTTGAACCGGCACTCCACTGCTGAT 3244  
Db 3198 GCGGAGACATCTCTGATGTTTCACTAGTTAGTTAACTTACCCCTACACCACTCTCCAT 3257  
Qy 3245 TCCAAGTCTTCCCGCATGTTCTAGTCTGCTGATGAGTTGAACCGCACTCCACTGCTGAT 3244  
Db 3258 GCAGGAGACGCCACATGTTTGGTGCATGTTGCAAGGCACTCTGTTCCCTCAAGTAC 3317  
Qy 3305 TACACCTGAGTGTGATGCTTTAGTACAGATAATTTAGTGTATTTGCTCTTTTAAAGCTCC 3364  
Db 3318 TACACTGTTGTATGATGCTTTTGTAGTTAGTATCATTAGAGATCTAGAGAACTTAGCTCG 3377  
Qy 3365 TTCTTTTGAAGATGATATGATGAGAGAGTGTAGTAGTGAATTTAGTGTATTTTAAAGCTCC 3424  
Db 3378 TACTGTTAGATATGATTAAGTGTGATGAGGAAACCAATAGCAATTTACAGATTTGACTCG 3437  
Qy 3425 GTGTTCAAGGTCATATCTCTTTGTGGCAACCTTAATCAGGAGACTTTCCAGATCTA 3484  
Db 3438 GTGTTCAAGGTTCCATCTTTTGTGCGCGCCAAAGACTGTTGATTTCTGATATG 3497  
Qy 3485 CAGTTCTATTACGATGATGCTCCCTGTTAATAGTACTATACTTAAACAGATGATGCT 3544  
Db 3498 CAGTTTACTATGATTAAGTGTCTCCAGGCAACAGCACCATGATGAATAATTTTGTATGCT 3557  
Qy 3545 GTTACCATGAGGTACGTTGATTAATGTTGAAGGATTTGTTCTTGTGATTTTCC 3604  
Db 3558 GTTACCATGAGGTGACTGACATTTCAATGATGTTCAAGATTTGATGATTTGATGATGCT 3617  
Qy 3605 AAAAGTATTCGATGCCAAAGAGGTGAACCATGTCTAGAGCCAGTTTTGGTACCCGCG 3664  
Db 3618 AAGTCTGTGCTGCCCTTAAGGATCAATCAACCACTAATACCTATGTTAGTACGAGCGCG 3677  
Qy 3665 GCGGAACCGCCAAAGGCTGCGAGGACTACTCGAAAACTCTGGTTGCAATGATTAAGAAAT 3724  
Db 3678 GCAGAAATGCCACGCGCAGACTTGGACTATTGGAAAAATTTAGTGGCGATGATTAAGAGGAA 3737  
Qy 3725 TTCAAACGACAGACTGACGGGAGGATTCACATTTGAGGACCGCATCTGTTGTAGTA 3784  
Db 3738 TTTAAACGACCGGAGTGTCTGGCATCATTTGATATTGAAATTTGAAATTTGATGTTGTA 3797  
Qy 3785 GATAAGTTTTGTAGTCTTATTTTATAAAGGAAAAATAACAATAAATAATTTCTGGA 3844  
Db 3798 GATAAGTTTTGTAGTCTTATTTGCTTAAAGGAAAAAGAAAAACCAATAAATAATTTCT 3857  
Qy 3845 GTGATGACGAAGGATTCATGATGATGTTGGAAAAACAGGAAAGAGTACTATTGGAC 3904  
Db 3858 TTGTTTCAAGTAGAGTCTCTCAATAGATGTTTGAAGAAACAGGAAACAGTAAACATAGGC 3917





1040 CAACTAGGCTGTAGTATGATCAGTTCTATGAGCGGATGGAAGACCGCTTTGCTTACAAG 1099  
1038 CATAAAAGTGTAGATGAGCAGTTTATATCTGCAATGGAAGACGATGCGCATTCANAA 1097  
1100 AAAACCTTTGGCCATGTTCAACACATGAAAGAGCAATCTTTAGAGACACGGCTTCGGTTAAC 1159  
1098 AAGACTCTTGGCAATGTGCACAGCGAGAGATCCTCTTGGAGATTCAATCATCACTCAAT 1157  
1160 TTTTGGTTCCCTAAGATGAAGGAACATGGTGTATGATACCGCTGTTTGGAGGTTCTATTACC 1219  
1158 TACTGGTTTCCCAAAATGAGGGAATGCTCATCGTACCATTCATTCGACATTTCTTTGGAG 1217  
1220 AGCAAAAGATGACAAGGAGTGAAGTCAATGTTTAAATCGTGACTTTCGTTTACACAGTCTT 1279  
1218 ACTAGTAGAGGACGCGCAAGGAAGTCTTAGTGTCCAAGGATTTTCGTTTACAGTCTT 1277  
1280 AATCAATCAGAACATATCAAGCCAAAGCGTTAACTTACAGAACGTAATATTCTTTCTGTG 1339  
1278 AACCAATTCGAACATACAGCGCAAGAGCTCTTACATACGCAATGTTTGTCTTTGTCTGTC 1337  
1340 GAGTCTATAAGATCCCGGTGATTAATCAATGGTGTACTGCTAGGTCTGAATGGATGTA 1399  
1338 GAATCGATTGATCGAGGGAATCAATTAAGCGTGTGACAGCGAGGTCCGAATGGATGTG 1397  
1400 GATTAAGCAATTTCTCAACCTTGTCAATGATCTTCTTCTGACAGCTTAAGCTGGCTGG 1459  
1398 GACAAATCTTTGTTACAAATCCTTGTCCATGACGTTTACCTGCACTATCAAGCTTTGCCGTT 1457  
1460 CTTCAAGACGATATAGTAATGGGAAGTTTCGGTGTGGATAAGACCACTTCTGAACTT 1519  
1458 CTAAGAGATGACTTACTGATTAGCAGTTTATGTTAGTCTCGTTTGGAAACCGTGTGCCAGAT 1517  
1520 ATTTGGGATGAGGTGGGCAAAATTTTGGAAACGTTTTTCCCACTATCAAGAGAGATTG 1579  
1518 GTGTGGGATGAGATTTTCGTCGGCTTTGGGAACGCAATTTTCCCTCGTGAAGAGAGGCTC 1577  
1580 GTGAGCAGGAAATTTCTGGATGTAAGTGAGAACTCTCTGAAGATCAAGATCCCAAGATCTG 1639  
1578 TTGAACAGGAACTTATCAGAGTGGCAGGCGACGATTTAGAGATCAGGGTGCCTGATCTA 1637  
1640 TATGTCACTGGAAGACAGGTTCTGTAGCTGAATACACCAAGTCTGAGGAGTTACCGCAT 1699  
1638 TATGTGACTTCCACGACAGATTAGTACTGATGATCAAGGCTCTGTGGACATGCTCTGG 1697  
1700 CTAGATATCAAGAAAGGACTTAGAAGAGCTGAGCAAAATGTACGACGGCTTATCAGAAATTA 1759  
1698 CTTGACATTAGAAGAAAGATGGAAGAAAGGAAGTGTATCAATGCACTTTTCAGAGTTA 1757  
1760 TCTATCCTTAAGGCTGCTGATAATTTTCGATATCGGAAGTTCAAGACATGTGCAAGCT 1819  
1758 TCGGTGTTAAGGGAGTCTGACAAATTCGATGTTGATGTTTTTCCCGAGATGTGCCAATCT 1817  
1820 TTAGATGTTAGTCTGTGATGCGGACGAGTAAATCGTTGAGTGGCGGAGAGATAGAGC 1879  
1818 TTGGAAGTTGACCAATGACGCGAGCGAGGTTATAGTTCGGGTCATGAGCAATGAGAGC 1877  
1880 GGTTTAACTCTTACTTTTGAATGAAGCCAAACCGAGGAGATGTGGCTAAAGGCTCT-----T 1933  
1878 GGTCTGACTCTCACAATTTGAACGACCTACTGAGCGGAATGTTGCGCTAGCTTTTACAGGAT 1937  
1934 AAAAGCAGGCTCTGAGCGCGTGGTATGTTCTTGAACCGACATCCGAGAGGTTGAACGTA 1993  
1938 CAAGAGAGGCTTCAGAAGGTGCTTTTGGTATTACCTCAGAGAAAGTTGAAGAACCGTCC 1997  
1994 AATAAAATTTCTATTGCTGAGAAAGGAGATTGCTGTGCTGACAGAAAGTCATGGTTTG 2053  
1998 ATGAAGGTTTCGATGCGCAGAGGAGATTACAATTAGCTGTCTTCTGTTGAGATCATCCG 2057  
2054 ACGAATGCTAATCTTAGAGCACAGGAGTTGGAGTCCCTCAACGATTTTCCATAAGGCTTGC 2113  
2058 GAGTCGCTCTATTCTAAGAACGAGGAGATAGAGTCTTTTAGAGCAGTTTTCATATGSCAACG 2117  
2114 GTGGATAGTGTGATTACAAGCAAAATGCGATCGGTTGTCTACACTGGCTCACTCAAAAGTT 2173

2118 GCAGATTTCGTTAAATTCGTAAGCAGATGAGCTCGAATGTGTACCGGTCCGATTAAGTT 2177  
2174 CAACAAATGAAGAACTATGTGGACAGTTTGGCAGCTTCGTTGTCCGCCACTGTATCAAAAT 2233  
2178 CAGCAATGAAGAACTTTATCGATAGCTGTGTAGCATCACTATCTGTCTGCGGTGTGCAAT 2237  
2234 CTATGCAAGTCACTAAAGGATGAAGTCGGGTATGATTTCTGATTCAGGGGAGAAAGTTGGT 2293  
2238 CTGTCGAAGTCTCTCAAGATACAGCTGCTATTGACCTTGAACCCGCTCAAAAGTTTGGG 2297  
2294 GTTTGGGATGTCACTTTGAAAGTGTCTCTCAAACTCGCGGCCAAAGGTCATTCAATGG 2353  
2298 GTCTTGGATGTGCAATCTAGGAGTGTGTTAATCAACCAAGCGGCAAGAGTCATGCAATGG 2357  
2354 GGAGTTCTCTCGGATTAACAAGGGGAAATGTTTACTGCACTTCTATCTTATGAAGGAGAT 2413  
2358 GGTGTTGTTGAAACCCACGCGAGGAAGTATCATGTGGCGCTTTTGGAAATATGATGAGCAG 2417  
2414 AGAATGTGTACTGAGACGCACTGAGGAGGGTGTCTGTATCATCTGTATCAATGGTATAT 2473  
2418 GGTGTGTGATCATGCGATGATTTGGAGAGAGTAGTCTGTCACTCTGTGATCTGTGTTTAT 2477  
2474 TCTGATATTCGAAGCTCCAAATCTCAGGAAACAAATGAGAGACGGTGAACCCCAAGAA 2533  
2478 TCCGACATGGCAAACTCAGAACTCTGCGCAGACTGCTTCGAAACGGGAAACCGCATGTC 2537  
2534 CCTACTCAAGAGATGGTACTTTGTGGATGGGTGCTCGTTGTGTGAAAGTACAAAGGAGAT 2593  
2538 AGTAGCGCAAGGTGTTCTTGTGACGAGTTCGGGCTGTGGGAAACCAAGAAATTT 2597  
2594 TTTGAAAGATTTGATCTTGTATGAGGATTTGATCTTGGTTCCTGGAAAACAAGCTGTCTCT 2653  
2598 CTTTCCAGGGTTAATTTTGTATGAAGATCTAATTTTAGTACCTGGGAAACAAAGCCGAA 2657  
2654 ATGATCAGAGAGAGGGCTAATCTGATGATGAAGAGCCAAATGAGCAATGTGAGA 2713  
2658 ATGATCAGAGAGAGGTGGAATTTCTCAGGGAATTTATGTGGCCACAGAGGACAACTTANA 2717  
2714 ACGGTAGATTCACCTTCTAATGTCAT-----CCAAAAACCGCATCAACAAGAGGCTT 2764  
2718 ACCGTGATTTCTTATGATGAAATTTTGGGAAAGCACAGCTGTCTGAGTTCAAGAGGTTA 2777  
2765 TTTATGATGAAGGTTGATGCTGCACACCGGTTGTGTAACTTCTCTGCTGGTCTTATCTCT 2824  
2778 TTTCAATGATGAAGGTTGATGTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2837  
2825 GGTTCGACATCGCATACATTTACGGAGATACACAGAGATTTCTTTTCAATTAACAGAGTT 2884  
2838 TTGTGCGAAATTTGATATGTTTACGGAGACACACAGCAGATTTCCATATCATCATAGATTT 2897  
2885 CAGAAATTTCCCGTATCCCAACATTTTGAAGCTGCAAGTGGATGAAGTTGAGATGAGG 2944  
2898 TCAGGATTTCCCGTATCCCGCCCAATTTTGGCAAAATTTGGAAGTTGACGAGGTGAGACACG 2957  
2945 AGGACCACTGAGATGCCAGGATGTGAATTTTTTCTCAATCGAATGGAAGTGAAGGA 3004  
2958 AGAATCTCTCTCCGTTTCCAGCGCATGTCAACATTTCTGAAACAGGAGATATGAGGCG 3017  
3005 GCGGTGACCAACCTTCAACTGTACAGATCGGTCTCATCTGAGATGATAGGCGGTAAG 3064  
3018 TTTGTGATGAGCATTCTTCTCGTTTAAAGTCTCTTTCGCGAGGAGATGGTTCGGCGGAGCC 3077  
3065 GGAGTACTAAACAGTGTTCCAAAACCACTAAAGGGGAAATTTGTAACCTTTTCACTCAGGCT 3124  
3078 GCGGTGATCAATCCGATCTCAAAACCTTGCATGCGAAGATCTTGTACTTTTACCAATCG 3137  
3125 GATAAATTTGAGTTAGAGGAGAGGGCTATAAGAAATGTGAACACCGTTTCATGAGATCCAA 3184  
3138 GATAAAGAGCTCTCTCTTTCAAGAGGGTATTTCAGATGTTTCACTGTGTCATGAAGTGAA 3197  
3185 GGAGAAACCTTTGAAGATGTCGCTGTCAGATTGAGCGCAACTCCCACTGCTCTGATT 3244

Db 3198 GCGAGACATACCTGATGTTTCACTAGTTAGGTTAAACCCCTACACCAAGTCTCCATCAT 3257  
Qy  
3245 TCCAAGTCTTCCCGCAGTCTTCTAGTCTGCTCTGACTAGACACACAAAGAGCTTCAATAT 3304  
Db  
3258 GCAGGAGACAGCCACATGTTTGGTGCATTTGTCAAGGCACACCTGTTCTCAAGTAC 3317  
Qy  
3305 TACACGCTAGTGTAGATCCCTTTAGTACAGATAATTTAGTGAATTTGCTCTTTTAAAGTCC 3364  
Db  
3318 TACACTGTGTATGATGATCCCTTTAGTGTATCATTTAGATCTTAGAATACTTAGCTCG 3377  
Qy  
3365 TTCTCTTTTAGAATATATATGATAGACAGCTAGTAGATCAATTTACAGATGATGCA 3424  
Db  
3378 TACTTTGTAGATATGATAAGGTGATGATGATGATGATGATGATGATGATGATGATGATG 3437  
Qy  
3425 GTGTTCAAAGGTGATATCTCTTTGTGCAACACCTTAATCAGGAGACTTTCCAGATCTA 3484  
Db  
3438 GTGTTCAAAGGTGATATCTCTTTGTGCAACACCTTAATCAGGAGACTTTCCAGATCTA 3497  
Qy  
3485 CAGTTCTATTTAGATGATGCTCCCTGCTGATGATGATGATGATGATGATGATGATGATG 3544  
Db  
3498 CAGTTTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3557  
Qy  
3545 GTTACCATGAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3604  
Db  
3558 GTTACCATGAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3617  
Qy  
3605 AAGATATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3664  
Db  
3618 AAGTCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3677  
Qy  
3665 GGGGAACCGCAAGGCTGCGAGTCTCTGAAATCTGTTGCAATGATTAAGAAAT 3724  
Db  
3678 GCAGAAATGCGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3737  
Qy  
3725 TTCAACGCAACAGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3784  
Db  
3738 TTCAACGCAACAGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3797  
Qy  
3785 GATAGTCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3844  
Db  
3798 GATAGTCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3857  
Qy  
3845 GTGATGACGAGGATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3904  
Db  
3858 TTGTTCACTAGAGTCTCTCAATAGATGATGATGATGATGATGATGATGATGATGATG 3917  
Qy  
3905 GACTGGCTTAACTCAATTTTACAGATCTGCGCGCATCGATCAGTACAGACATGATC 3964  
Db  
3918 CAGCTCGCAGATTTGATTTTGTAGATTTGCCAGCAGTTGATCAGTACACATGAT 3977  
Qy  
3965 RAGGCTCAACCAAAACAGAAATTTGACCTTTCAATTTCAATTTCAATTTCAATTTCAAT 4024  
Db  
3978 AAGACAAACCAACCAAAATTTGACCTTTCAATTTCAATTTCAATTTCAATTTCAATTT 4037  
Qy  
4025 ACAATTTGCTTACCATTTCAAGCAGATCAACGGTATTTTGGCCGGTTTCTCA --- GAGCTT 4081  
Db  
4038 ACGATTTGCTTACCATTTCAAAAGATCAATGCAATTTTGGCCGGTTTGTAGTACGCTT 4097  
Qy  
4082 ACAAGTCTGCTCGAGCAGTCTGATTTCTAGAGTCTTCTTCTTCTTCTTCTTCTTCTTCT 4141  
Db  
4098 ACTAGGCAATTTCTGACAGTGTGATTCGAGCAGATTTTGTGTTTCTCAAGAAAGCA 4157  
Qy  
4142 CAGAACAGATTTCAAGAAATTTTCTCGGATCTCGACTCGCAGCTTCTTCTTCTTCTTCTT 4201  
Db  
4158 CAGGCGCAGATTTGAGGATTTCTTCTCGGATCTCGAGTCTGCGAGTCTGCGGATGATGCT 4217  
Qy  
4202 GAATCGGATTTCTTAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 4261  
Db  
4218 GAGCTGGATATCAAAATACGACAAATCTCAGATGATGATGATGATGATGATGATGATG 4277  
Qy  
4262 GAATATGAAAGATTTGGTCTCAATGATTTTGGCCGAGTGTGGAAACAAAGGCAAC 4321  
Db  
4278 GAGATCTGGGAGATTTGGGTTTGAAGACTTCTTGGGAGAGTCTTGGGAGAGTCTTGGGAG 4337

Qy 4322 AGGAAACAACTTTGAAGGATTTACATTTCTGCTGAATCAAGACATGCTCTGTGTATCAAGG 4381  
Db  
4338 AGAAGACCACTCTCAAGGATTTATACCGCAGGTTATAAANAATGCTCTGTATCAAGA 4397  
Qy  
4382 AAAAGCGGTGATGATGATCTTTTCATCGGCAATCTGTTTATAAGAGCTGCTGCTGGT 4441  
Db  
4398 AAGAGCGGACGCTCACGAGTTCAATTTGAACACATCTGATCATTTGCTGATGCTTTGGCC 4457  
Qy  
4442 TCAATGTTACCGATGGAAGGTCATAAAGGTGCTTTTGTGAGAGAGTCTCGTTTGTG 4501  
Db  
4458 TCGATGCTCCGATGGAAGAAATAAATAAGGAGCTTTTTCGGTGAGATGATGCTGCTG 4517  
Qy  
4502 TATTTTCCAAAGGTTTGGATTTCCCTGACATTCAGTCAATGCTGCTAATCTCATGTGGA 4561  
Db  
4518 TACTTTCCAAAGGTTTGTGATTTCCGATGTCACACATCCCGCAATCTTATGTGGA 4577  
Qy  
4562 TTTGAGGCCAACTGTATAGAGAGTACCGTTTACTTTTGTGTTGTGATATCATATACAC 4621  
Db  
4578 TTTGAAGCAAACTGTTTAAAAAAACAGTATGGATACCTTTTCCGGAAGATATGTAATAC 4637  
Qy  
4622 CATGATAGGAGCAATAGTGTATTTATGATCTTTGAAGTTGATCTCCAAACTTTGGGCA 4681  
Db  
4638 CACGACAGAGATGCTATTTGATTTACATCCCTTAAAGTTGATCTCGAACTTTGTGCT 4697  
Qy  
4682 AAAATATCAAGGATTTATGATCACTTAGAAGATTAAGGTTGCTTTTGTGCGATGTTGCT 4741  
Db  
4698 AAACATCAAGGATTTGGAAACCTTGGAGAGTTTCAAGAGTCTCTTTGTGATGTTGCT 4757  
Qy  
4742 TGTTCGCTCGAATACTGTTAGGCTTTCCGAGCTGACGAGCTATCAAGAGGTT 4801  
Db  
4758 GTTTTCGTT --- GAAACAAATTTGCGTATTTACACAGTTTGGACGAGCTGATGGAGGTT 4814  
Qy  
4802 CATAAAACCGGATTTGATGTTTCTGTTTAAATTTGTTTAAACAAATTTTGTGAT 4861  
Db  
4815 CATTAAGCCGCTTCCAGGTTTCTGTTTATTAAGTCTGTTGAGATTTTGTCTGAT 4874  
Qy  
4862 AAATTTTATTTAGAACTTTGTTTAAATGGCTGTTAGTCTCAGAGATCTGTCAAAAT 4921  
Db  
4875 AAAGTTCTTTTAGAAGTTTCTTTATAGTGGCTCTAGTTTGTAAAGGAAAGTGAATAT 4934  
Qy  
4922 TAGCGATTTCAATCTTTTCGAAACAGGATGATATCTCCGCAATTCATGACTTAAGGT 4981  
Db  
4935 CAATGAGTTTATGCTGCTCACTGACAAATAAGGAGAGATCTTACCGTCAATTTACCCCTGT 4994  
Qy  
4982 CAAGAGTTTAAAGATATCGACTGTGGCAAGATTTAGCTGTTAAGAAATGATGCTTTTC 5041  
Db  
4995 AAAGGTTTATGTTTCCAAAGTTGATAAATAATGTTTCTATGAGATGATGATGATGCTC 5054  
Qy  
5042 TGATGATGATTTCTTAAAGGTTTAAAGTATGTTAAGAAAGGTTATGTTGCTTAGCTGA 5101  
Db  
5055 AGAGTGAACTTTCTTAAAGGATTTAAGCTTTATTTGATAGTGGATACGCTCTGTTAGCCGG 5114  
Qy  
5102 TTTGTTAGTGTCTGGGAGTGAATCTCCCGATAACTGCGGTGTTGTTGCTGAGTGTGTTG 5161  
Db  
5115 TTTGTTGCTGCTGCGGAGTGAATCTGCTGCAATTTGAGAGAGGTTGAGAGGTTG 5174  
Qy  
5162 TATTGATAGAGAAATGAAGAGGTTAAGAAAGCAACGCTGGGTGCTGATCAGCCCC 5221  
Db  
5175 TCTGTTGACAAAGGATGGAAGAGCGGACGAGGCGACTCTCGGATCTTACTACACAGC 5234  
Qy  
5222 TGCTTGGAAAGAAATTTTCTTTTAACTAATCCCTAATTTTCAATAACATCCGAGGA 5281  
Db  
5235 AGCTCAAGAAAGAAATTTTCAAGTTCAGGTCGTTCCCAATTTATGCTATTAACCCAGGA 5294  
Qy  
5282 TGCTGAGAGCAACCGCTGCGCAAGTGTAGTGAATATAAGAGGTTGGCTTATGGAAAGAG 5341  
Db  
5295 CCGGATGAAGAAAGCTGCTGCAAGTTTATGTTAATTAAGAAATGTAAGAGTGTGAGCGG 5354  
Qy  
5342 ATACTGCTCTTTATCTTTGGAGTCTGTTTCAATTTGTTGTGATGATCAATAAATAATGTAAG 5401  
Db  
5355 TTTCTGCTCGCTTTCTCTGGAGTTTGTGCTCGGTGTTGTTTATAGAAATAATATAA 5414

```
Oy 5402 AAAAGGTTTGGGAAACGTATTTTGTAGTGTGACAGCGGCTCGCAATTGAACTCACTGA 5461
Db 5415 ATTAGGTTTGAGAGAGAAGATTACAAACGTGAGAGACGGAGGCCCAATGGAACTTACAGA 5474
Oy 5462 AAAGGTTGTTGAGAGGTTGTTGAGTGAAGTACCAATGGCTGTGAATCTCGAAAGGTTTC- 5520
Db 5475 AGAAGTGGTGTGATGATGTTTCATGGAAGATGTCCTTATGTCGATCAGGCTTGCAAGTTTCG 5534
Oy 5521 -----CGAAAAACAAAAAGAAAGTGTAGGTAATATGTTTAATAT----- 5561
Db 5535 ATCTCGAACCGAAAAAGAGTGATGTCGCAAGGGAATAATAGTAGTAATGATCGTGC 5594
Oy 5562 -----AAGAAATAATACATGCTGTGAAGAGGTTTAAATTTGAGAAATGAGGA 5614
Db 5595 AGTCCGAACAAGAACTATAGAAATGTTAAGGATTTTGGAGGAATGAGTTTAAAAAGAA 5654
Oy 5615 TAATGTAAGTGATGACAGAGT-----CTATCGCTCATCGAGTAGTGTTTAAATCAAT 5665
Db 5655 TAAATTAATCGATGATGATTCGGAGGCTACTGTGCGCAATCGATTCGTTTAAAT 5711
Oy 5666 ATGCTTTATACAACTCTCCGAGGCAATTTGTTTACTTTATCTTCGCTTACGCGAGAT 5725
Db 5712 ATGCTTTACATGATCACTACTCTCCATCTCAGTTCGTGTTCTGTCTATCAGCGTGGCGGAC 5771
Oy 5726 CCTGTCAGCTGATCAATCTGTGTACAAATGCGATGGTGTACCACTGTTTCAACCAACAA 5785
Db 5772 CCAATAGAGTTTAAATTAATTTATGTAATGTAATGCTTTAGGAAATCAGTTTCAAAACAA 5831
Oy 5786 GCTAGGACAAACAGTCCAAACAGCAATTTGCGGATGCTGGAACCTGTGCTAGTATGACA 5845
Db 5832 GCTCGAATCTGCTTCAAGACAAATTCGTGAGGTGTGGAACCTTCCCAACAAGTAAT 5891
Oy 5846 GTGAGATTTCTGCTGATCGGATTTCTATGTGTATAGATATAATTCGACGCTTGATCCGTTG 5905
Db 5892 GTTAGGTTTCTGACAGTGACTTTAAAGGTGTACAGGTACAAATGCGGTATTAGACCCGCTA 5951
Oy 5906 ATCAGCGGTTTAAATAGCTTTGTATAGTACTAGAAATAGAAATAGAGGTTGATAATCAA 5965
Db 5952 GTCAAGCACTGTTAGGTGCAATTCGACACTAGAAATAGAAATAGAGGTTTGAAGATCAG 6011
Oy 5966 CCGCACCGAATCTACTGAAATCGTTAAACCGACTCAGAGGGTAGAGATGCTACTGTGA 6025
Db 6012 GCGAACCCGACGCTGCGGAAACGTTAGATGCTACTCGTAGAGTAGACGACGCAACGGTG 6071
Oy 6026 GCTATAAGGGCTTCAATCAATAATTTGCTTAATGAACCTGTTCTGTGGAACCTGGCATGTT 6085
Db 6072 GCCATAAGGAGCGGATAATAATTTAATAGTAGAATTTGATCAGAGGAACCGGATCTTAT 6131
Oy 6086 AATCAGCAGGCTTTGAGACTGTAGTGACTTGTCTGGACCAACATCCCGGCTACTTAG 6145
Db 6132 AATCGGAGCTTTTCGAGAGCTCTTCTGGTTTGGTTTGGACCTCTGGTCTCTGCAACT--- 6188
Oy 6146 CTATTGTTGTGAGATTTCTTAAATAAGTCGCTGAAGACTTAAATTCAGGTTGGCTGA 6205
Db 6189 -TGAGGTAGTCAAGATGATTAATAATGCGATTTGTGCTGCTAATCA- -CGTGGTGGC 6246
Oy 6206 TACCAAAATCAGCAGTGTGTTGCTCCACTTAATAATATAACGATTTGTGCATATCTGGATCC 6265
Db 6247 TACGATAACGATAGTGTGTTTCCCTCCACTTAATAATCGAAGGTTGT- -GTCTTGGATCGC 6305
Oy 6266 AACAGTTAAACATGTGATGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 6325
Db 6306 GCGGGTCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 6365
Oy 6326 AATCTCTCCCTAAACCGCGGTAGGGGCCCA 6355
Db 6366 AATCCCCCGGTACCCCCCGGTAGGGGCCCA 6395
```

RESULT 7

US-09-259-741-2

; Sequence 2, Application US/09259741

; Patent No. 6033895

```
GENERAL INFORMATION:
APPLICANT: GARGER, STEPHEN
APPLICANT: HOLTZ, R. BARRY
APPLICANT: MCCULLOCH, MICHAEL
APPLICANT: TURPEN, THOMAS
TITLE OF INVENTION: A PROCESS FOR ISOLATING AND
TITLE OF INVENTION: PURIFYING VIRUSES SOLUBLE PROTEINS AND PEPTIDES FROM PLANT
TITLE OF INVENTION: SOURCES
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
STREET: 1299 Pennsylvania Avenue N.W.
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/259,741
FILING DATE: February 25, 1999
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/037,751
FILING DATE: March 10, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Halluin, Albert P
REGISTRATION NUMBER: 25,277
REFERENCE/DOCKET NUMBER: 00801.0140.US01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-463-8100
TELEFAX: 650-463-8400
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 6439 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: Genomic RNA
US-09-259-741-2
```

Query Match 38.6%; Score 2455.4; DB 3; Length 6439;

Best Local Similarity 44.8%; Pred. No. 0;

Matches 2772; Conservative 1147; Mismatches 2206; Indels 60; Gaps 8;

```
Oy 20 ACAACAACAATTTAAACAAACAAACATATTACAAACAAACAAACAAACAAACAAACAAACAAAC 79
Db 18 ACCAACCAACAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 77
Oy 80 ATACAATCTATATTAGCAACGCCCTTCTTGAAGCGTGAGTGTGTAAACAAACACTCTCGTT 139
Db 78 ACACAGACAGCAUACCAUACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 137
Oy 140 AATGACCTTGCAGAAAGCGCATGTAGATACGCGCGCTGGAGAAATTTAAACGCGCGAC 199
Db 138 AAUGAUCAGCAAGCGGUCUUAACGACACACGCGGUGAAGAGUUAUUAACGCGGAC 197
Oy 200 CGTAGACAAAGGTCAATCTTTTCCAAAACATATTAGCGAAGAGCAAAACGCTTCTAGTCTCC 259
Db 198 CGCAGGCCCAAGGUGAACUUUUAACAAAGAAUAAAGCGAGGAGCAGACGCGUUAUUGCUACC 257
Oy 260 AACCGGTACCGGAGTTCAGATTACCTTTTATATCTCAAAATCTCAAAATCTCAAAATCTCAAA 319
Db 258 CGGCGGUUACGAGAAUUAUUAACGAGAAUUAUUAACGAGAAUUAUUAACGAGAAUUAUUAACG 317
Oy 320 GCTGAGGTTTGAGACGATTAGAAATGGAATATCTGATGCTACAAAGTTTCCCTATGGATCG 379
Db 318 GCAGGUGAUGCGUUAUUAUUAACGAGAAUUAUUAACGAGAAUUAUUAACGAGAAUUAUUAACG 377
```



























QY	1100	AAACCTCTGGCCATGTTTCAACACTGAAAGAGCAATCTTTAGAGACACGCGTTCGGTTAAC	1159
DB	1098	AGACUCUUGCAUUGUGCAACAGCGAGAGAUCCUUGGAGAUCAUCAUCAGUCAU	1157
QY	1160	TTTTTGGTTCCTTAAGATGAAGGACATGGTGAATAGTACCCTGTTGTGAGGGTTCATTATACC	1219
DB	1158	UACUGGUUCCCAAAUAGAGGAUUGGCAUCUGUACCAUAUAUUGGACAUUUUCUUGGAG	1217
QY	1220	AGCAAAAGATGACAAAGGAGTGAAGTTCATTGTTTAATTCGTGACTTTCGTTTACACAGTGGT	1279
DB	1218	ACUAGUAGAGGACCGCAAGAGUAGUUGUCCAGGAUUUCUGUUAUACAGUGCUU	1277
QY	1280	AATCATATCAGAACATATCAAGCCAAAGCGTTAACTTACAGAACGTATATCTTTTCGTG	1339
DB	1278	AACCACAUUCGAACAUAACAGCGCAAGTCUUAUCAUACGCAAAUGUUUGUCCUUGUC	1337
QY	1340	GAGTCTATAGATCCCGCGTGATAATCAATGGTGTACTGCTAGTGTCTGAATGGGATGTA	1399
DB	1338	GAUCGAUUCGAUCGAGGGUUAUUAUACGUGUGACAGCGAGUUCGUAUGGAUGUG	1397
QY	1400	GATAAAGCAATCTTCAACCCCTTGTCAATGACTTTCTTTTCGAGACTAAGCTGGTGGG	1459
DB	1398	GACAAUCUUGUUAACAUCUUGUCUUGCAUGACGUUUUACUGCAUAUAAGCTUUGCCGU	1457
QY	1460	CTTCAAGACGATATAGTAATGGGAAGTTCGGTGTCTGGATAAGACCACTTCTGAACTT	1519
DB	1458	CUAAGGAUACUUAUGAUAGCAUUUAUUCUGUUCGUGUUGAAACGGUGUCCAGCAU	1517
QY	1520	ATTTGGGATGAGGTGGGCAAAATTTTTTGGAAACGTTTTTCCCACTATCAAAAGAGATTG	1579
DB	1518	GUGUGGAUGAGAUUUCUGUGGUUUGGAAACGCAUUCUCCUGGAAAGAGAGACUC	1577
QY	1580	GTGAGCAGGAAATCTGGATGTAAAGTGAGAAATGCTTGAAGATCAAGATCCAGATCTG	1639
DB	1578	UUGAACAGGAAACUUUAUCAGAGGCGAGGCGACGCAUUAAGAUAGGAGGUGCCUGCAUUA	1637
QY	1640	TATGTCATCATGAAAGACAGGTTCTGATGCTGAATATACACCAAGTCTGAGGAGTTTACCGCAT	1699
DB	1638	UAUGUGACCUCCAGCAGAUUAGUGACUGAUACAAGGCCUCUGUGGACAUUGCCUGCG	1697
QY	1700	CTAGATATCAAGAAGGACTTAGAAGAGCTGAGCAATGTACGACGCGTTATCAGAAATTA	1759
DB	1698	CUGACAUUAGGAAGAAGUAGGAAGAAACCGGAAGUGAUGUACAUGCAUUUCAGAGUUA	1757
QY	1760	TCTATCCTTAAGCGTGTGATAATTTCCGATATCCGGAAGTTCAAAGACATGTGCAAGGCT	1819
DB	1758	UCGGUUAUAGGGAGUCUGACAAUUCGAUGUUGAUUUUUUCCAGAUUGGCCAAUCU	1817
QY	1820	TTAGATGTTAGTCTGTAGTGTGGCAGCAGGAGTAATCGTTGAGTGGCCGAGAGATAGAGC	1879
DB	1818	UUGGAAUGUAGCCCAUAGCGGACGAGGAGGUUAUUGUCGCGGUCAUGACCAUAGAGAGC	1877
QY	1880	GGTTTAACCTCTACTTTTGAATGAAGCCAGGAGAGATGTGGCTAAGGCTCT-----T	1933
DB	1878	GGUCUGACUCUCAUUAUGACGACCUACUGAGGCGGAUUGUGCGUAGCUUAUACAGGAU	1937
QY	1934	AAAAGCACGGCGTCTGAGGCCGTGGTATGTCTTGAAACCGACATCCGGAAGAGGTGAACGTA	1993
DB	1938	CAAGAGAAGGCUUCAGAGAAGGUGCAUUGGUAUUAUCCUACAGAGAAGTUGAAGAACCGUCC	1997
QY	1994	AATAAATTTCTATTGCTGAGAAAGGGAGATTGCTGTGTGTGCAGAAAGTCATGTTGTG	2053
DB	1998	AUGAAGGUUCGAUGGCCAGAGGAGAGUUAUCAAUUGCUGUCUUGUGGAGAUCAUCCG	2057
QY	2054	ACGAATGCTAACTTAGAGCAACGAGGTTGGAGTCCCTCAAAGATTTTCATAAGGCTTGC	2113
DB	2058	GAGUCGCUUAUUCUAAAGAACGAGGAGAUAGAGUUAUUGAGCAGAGUUAUUAUGGCCAGC	2117
QY	2114	GTGGATGTGTGATTACAAAGCAAAATGTCATCGGTTGTCTACACTGGCTCACTCAAAAGTT	2173
DB	2118	GCAGAUCCGUUAUUCGUAGACGAUAGAGUCUGAUUGUUGUACAACGGUCCGAUUAAGUU	2177
QY	2174	CAACAAATGAAGAACTATGTGGACAGTTTGGCAGCTTTCGTTGTCGCGCACTGTATCAAT	2233

[illegible]









Db 2778 UUAUUGAUGAGGUGUGAUGCAUACUGGUGUUAUUAUUGGCGGAUGUCA 2837  
Qy 2825 GGTTCGCAATGCGCATACTTACGAGATACACAGCAGATCTCTTCAITAAACAGAGTT 2884  
Db 2838 UUGUGGAAAUUGCAUUGUUAUCGAGACACACAGCAGAUUCCAUACAUAAGAGUU 2897  
Qy 2885 CAGAATTTCCGTATCCCAACATTTTCAGAGCTGCAAGTGGATGAATGGAGG 2944  
Db 2898 UCAGAUUCCGUAUCCCGCCCAUUGCCAAUUGGAUUGAGCGAGUGAGACGC 2957  
Qy 2945 AGGACACACTGAGATGCGCAGGTGATGTGAATTTTCTCAATCGAAGTACGAAGGA 3004  
Db 2958 AGAAUAUCUUCUGUUGCCAGCCGAGUUCACAUUAUCUGAACAGGAGAUUAGAGGC 3017  
Qy 3005 GCGGTGACAAACACTTTCAACTGTAACAAGATCGGTCTCATCTGAGATGATAGGCGGTAG 3064  
Db 3018 UUGUGAUGAGCACUUCUGGUAUAAAGUCUGUUGCGAGGAGUUGGCGGAGGCC 3077  
Qy 3065 GGAGTACTAAACAGTGTTCCTAAACACCTATAAGGGAATTTGTAATCTTCACTCAGGCT 3124  
Db 3078 GCGUGAUCAUCCGAGUCCUAAACCCUUGCAUGGCAGAGUCCUUAUCCCAUUG 3137  
Qy 3125 GATAAATTTGAGTTGAGAGAGAGGCTATAAGAAATGTGAACACACCGTTTCATGAGATCAA 3184  
Db 3138 GAUAAAGAGCUCUGCUUAAGAGGUAUUCAGAUUGUUCACACUGUGCAUGAAGUGCAA 3197  
Qy 3185 GGAGAAACCTTTGAGATGTGCTGCTGCTGATTCAGCGCACTCCACTGACTCTGATT 3244  
Db 3198 GCGGAGACAUACUCUGAUUGUUCACUAGUUAAGUUAUCCUACACAGUCCUUAUUC 3257  
Qy 3245 TCCAACTCTTCCCGCATGTTCTAGTCTGCTGACTAGACACACAAAGAGCTTCAATAT 3304  
Db 3258 GCAGGAGACAGCCCAUUGUUGGUGCAUUGUCAAGGACACACUGUUGCUCUAAAGUAC 3317  
Qy 3305 TACACGATGTTAGATCTTTAGTACAGATAATTTAGTGTGATTTGCTCTTTTAACTCC 3364  
Db 3318 UACACUGUUAUGAUGAUCUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 3377  
Qy 3365 TTCTTTTGAATGTATGTAGAGAGGAGTGTAGATAGCAATTTACAGATGTATGCA 3424  
Db 3378 UACUGUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 3437  
Qy 3425 GTGTTCAAGGTCTAATCTCTTTGTGGCAACACCTTAATCAGGAGACTTTCAGATCTA 3484  
Db 3438 GUGUCAAAGUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 3497  
Qy 3485 CAGTTCTATTACATGTATGCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3544  
Db 3498 CAGUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 3557  
Qy 3545 GTTACCATGAGTTTACGTGATAATAGTCTTAAATGTGAAGGATGTGTTCTGATTTTCC 3604  
Db 3558 GUUACCAUGAGGUGUACGUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 3617  
Qy 3605 AAAAGTATCCGATCCGAAAGAGGTTGAACACCACTGTCTAGAGCCAGTTTTCGCTACCGG 3664  
Db 3618 AAGUCUGUGCGCCUUAAGGAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 3677  
Qy 3665 GCGGAACCGCAAGGCTGACGACTACTCGAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3724  
Db 3678 GCAGAAUUGCCACCCAGACUGGACAUUGAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 3737  
Qy 3725 TTCAACGACACACCTGACGCGGACGATTTGACATTTGAGAGCACCGCATCTGTTGTAGTA 3784  
Db 3738 UUAACGCAACCCGAGUUGUGGCAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 3797  
Qy 3785 GATAGTTTTTTGATGATTTTATTTAAAGAAAGAAATACACAAAATATTTCTGGA 3844  
Db 3798 GAUAAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 3857  
Qy 3845 GTGATGACGAAGGATTCATCATGATGATGTTGGAACACGAAAGAGTACTATTGGAC 3904  
Db 3858 UUGUUCAGUAGAGAGUCUCUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 3917

Qy 3905 GACTTGGCTAACTACAAATTTTACAGATCTGCGGCCCATCGATCAGTACAGACATGATC 3964  
Db 3918 CACUCGCGAGAUUUGAUUUGCCAGCAGUUGUAGUACAGACAGCAUGAUU 3977  
Qy 3965 AAGGCTCAACCAAAACAGAAATTTGGACCTTTCAATTCAGAAATGAAATACCTCTCTGCA 4024  
Db 3978 AAAGCAACCAAGCAAAAUUGGACACUUUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 4037  
Qy 4025 ACAATTTCTACCATTCGAGACAGATCAACGCTATTTTGGCCGCTTTCT---CAGAGCTT 4081  
Db 4038 ACAGUUGUGUACAUUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 4097  
Qy 4082 ACAAGGTTGCTGCTCGAGGCAATTTGATTTCTAAGATTTCTAAGATTTCTTTTCTTAGTAAACT 4141  
Db 4098 ACUAGGCAUUAUCUGGACAGUUGUACGAGAGAUUUAUUAUUAUUAUUAUUAUUAUUAUUA 4157  
Qy 4142 CCAGAACAGATTCAGAAATTTTCTCGGATCTCGATCTCGACCTCTCTATGATGTGTTA 4201  
Db 4158 CCAGCGCAGUUGAGGAGUUCUUGCGGAGUUCUCGACAGUCAGUUGCCGAGUUGUUG 4217  
Qy 4202 GAACTGATATTTCTAAGTATGATAGTCAAGAACGAGTTTCAATGTGTGTAGATAT 4261  
Db 4218 GAGCUGAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 4277  
Qy 4262 GAAATATGGAAGAAATTTGGGTCTCAATGAGTTTGTGCGGAAAGTGTGGAACAGGCGAC 4321  
Db 4278 GAGAUUCGCGAGAUUUGGUGUUAAGACUCUCUGGGAAGAUUUGGGAACAGAGGCAU 4337  
Qy 4322 AGGAAACCACTTTGAAGGATTTACATTTCTGGAATCAAGACATGCTCTGTGTATCAAGG 4381  
Db 4338 AGAAGAGCACCCUUAAGAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 4397  
Qy 4382 AAAACGCGTGTGTGACTTCTCATCGCAATCTGTTATAAGTATTAAGTGTGTTGCTTGGT 4441  
Db 4398 AAGAGCGGGACGUCACGACUUAUUGAAACACUGUGAUCAUUGUGCAUGUUGGCC 4457  
Qy 4442 TCAATGTTACCGATGGAAGGTCATAAAGGTCCTTTTGTGAGAGCATTCGCTTTTG 4501  
Db 4458 UCGAUGUUCGAGUUGGAGAAAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 4517  
Qy 4502 TATTTTCAAGGGTTGGATTTCCCTGACATTCAGTCAATGTGTATCTCATGTGGAAT 4561  
Db 4518 UACUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 4577  
Qy 4562 TTTGAGGCAAACTGTATAGAAAGGTPACGGTTACTTTTGTGTTAGTATCATATACAC 4621  
Db 4578 UUAAGCAAAACUGUUUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 4637  
Qy 4622 CATGATAGGAGCAATAGTGTATTATGATCTTTGAAGTTGATCTCCAAACTTTGGGCA 4681  
Db 4638 CACGACAGGAGUAGCAUUGUGUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 4697  
Qy 4682 AAAATATCAAGGATTTATGATCATTTAGAAGTTTAAGGGTGTCTTTGTGCGATGTTGCT 4741  
Db 4698 AAACACUACAGAUUUGGAAACACUUGGAGGAGUUCAGAAAGGUCUCUUAUUAUUAUUA 4757  
Qy 4742 TGTGCTCGGAAACTGTTGCTTTCGCGAGCTTTCGCGAGCTGAACGAGCTATCAAGAGTT 4801  
Db 4758 GUUUGUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 4814  
Qy 4802 CATAAACCGGATTTGATGTTGCTTTGTTTAAATTTGTTTAAACAAATTTTGTGTGAT 4861  
Db 4815 CAUAGACCGCCCGCCAGGUCUUGUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 4874  
Qy 4862 AAATTTTATTTAGAACTTTTAAATGCTGTTAGTCTCAGAGATACTGTCAAT 4921  
Db 4875 AAAGUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 4934  
Qy 4922 TAGCGATTCATTTGATCTTTTCGAAACAGGATCAGATCTTCCGCGATTCATGACTAAGT 4981  
Db 4935 CAUAGAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 4994







Db 2298 GUCUUGAUGUUGCAUCUAGGAAGUGGUUAUUAUCAAACCAACCGCCCAAGAGUACUGCAUGG 2357  
Qy 2354 GGAGTTGTCTCGATTTACAAGGGGAAATGTTTACTGCACTTCTATCTTATGAAAGAGAT 2413  
Db 2358 CGUGUUGUUGAAACCAACCGGAGGAUAUUAUGUGGCGCUUUUGAAUAUUAUGAGCAG 2417  
Qy 2414 AGAATGGTGACGTAGAGCGACGTGGAGGAGGCTGTGTATCATCTGTATACATGTATAT 2473  
Db 2418 GGUGUGUGAUGCAUGCGAUGAUGGAGAGAGUAGUGUGAGUGUGUGUGUGUGUUAU 2477  
Qy 2474 TCTGATATTGCAAGCTCCAAATCTGAGGAAACAACTGAGAGACGGTGAACCCACGAA 2533  
Db 2478 UCCGACAUCCGAAACUCAGACUCUGCGCAGUCUUGGAAACGGAGAACCGCAUGUC 2537  
Qy 2534 CTTACTGCAAGATGCTTGTGATGGGGTCCCTGGTGTGGAAAGTACAAAGAGAT 2593  
Db 2538 AGUAGCGCAAGUGUUGUUGUGACGAGUCCGGGUGUGGAAACCAAGAAAU 2597  
Qy 2594 TTTGAAAGATTTGATCTTGATGAGATTTGATCTTGGTCTCTGGAACAGCTGCTGCT 2653  
Db 2598 CUUCCAGGUGUUAUUUGAUGAUAUUAUUAGUACUUGGAAACGAGCGCGGAA 2657  
Qy 2654 ATGATCAGAAAGGCTTAATTCATCTGCACTGATAAGAGCCACAAATGCAATGTGAGA 2713  
Db 2658 AUGACAGAGAGUGCGAUAUCCUAGGGAUUAUUUGGCCACAGAGACAAAGUAAA 2717  
Qy 2714 ACGGTAGATTCATCTTAATGATCCAAACCCGCGATCACA-----CAAGAGGCTT 2764  
Db 2718 ACCGUUGAUUCUUAUGAUGAAUUUUGGAAACGACACGCGUCUGACUUAAGAGGUUA 2777  
Qy 2765 TTTATTGATGAAGGTTGATCTGCACACCGGTTGTGTAACTTCTCGTGTCTATCTCT 2824  
Db 2778 UUAUGAUGAAGGUGAUGUGUAUCUGUGUGUUAUUUUCUGUGCGCAUGUCA 2837  
Qy 2825 GGTTCGACATCGCATACATTTACCGAGATACACAGCAGATTCCTTTTCAATTAACAGATT 2884  
Db 2838 UUGUGCGAAUUGCAUUGUUUUCGAGACACACAGCAGAUUUUUCUUAUUAUUGG 2897  
Qy 2885 CAGAAATTCCTGATCCCAACATTTGAGAGCTGCAAGTGGATGAAGTGAAGTGAAGG 2944  
Db 2898 UCAGGAUUCGCUACCCCGCCCAUUAUUGCAAUUGAAGUUGACGAGUGGAGACACGC 2957  
Qy 2945 AGGACCACTGAGATGCGCAGTGTGATGTAATTTTCTTCAATCGAATGACGAAGA 3004  
Db 2958 AGAACUACUCCUGUGUCGACCGCAUGUCAACAUUAUCUGAAACAGGAGAUUAUGAGGC 3017  
Qy 3005 GCGGTGACAAACCACTTCAACTGTACACGATCGGTCTCATCTGAGATGATAGGCGGTAAG 3064  
Db 3018 UUUUGAUGACCAUUAUUGGUUAAAGUUGUUUUCGAGGAGUUGGUGCGCGAGCC 3077  
Qy 3065 GGAGTACTAAACAGTGTTCAAACCACTAAAGGGAAATTTGTAATTTTCACTCAGGCT 3124  
Db 3078 GCGUGAUAUCCGUAUCUCAAACCCUUGCAUGGCAAGAUCCUGACUUUAUCCAAUCG 3137  
Qy 3125 GATAAATTTGATTAGAGAGAGGCTTAAGAAATGTGAACACGGTTCATGAGATCOAA 3184  
Db 3138 GAUAAGAAGCUCUCUUAAGAGGUGUAUUCAGAUUGUCAACUGUGCAUGAAGUGCAA 3197  
Qy 3185 GGAGAAACCTTTGAAGATGTGCTGCTGATGAGGCAACTCCACTGACTCTGATT 3244  
Db 3198 GCGGAGACAUACUGAUGUUUUCUUAAGGUUAAAGUUGUUUUCGAGGAGUUGGUGCGGAGCC 3257  
Qy 3245 TCCAGTCTTCCCGCATGTTCTAGTCTGCTGACTAGACACACAAGAGCTTCAAAATAT 3304  
Db 3258 GCAGGAGACAGCCCAUGUUUUGGUGCAUUGUUAAGGACACCCUUGUCUUAAGUAC 3317  
Qy 3305 TACACCGTAGTGTAGATCTTTAGTACAGATAATAGTGTGTTCTTTTAAAGCTCC 3364  
Db 3318 UACACUGUUGUAUGAUGUUAUAGUAGUAUUAUAGAGUAUUAUAGAGAAACUUAAGCUCG 3377  
Qy 3365 TTCTTTTAGAAATCTATATGTAAGACGAGTGTAGTACGATTTACAGATGATGCA 3424  
Db 3378 UACUUUGAUAUGUAUAGGUGCAUGAGGAAACAAUAGCAUUAUUAAGAUUGACUCG 3437

Qy 3425 GTGTTCAAAGTGCATAATCTCTTTGTGGCAACACTAAATCAGGAGACTTTCCAGATCTA 3484  
Db 3438 GUGUUAAGGUGUCCAAUUAUUUGUUGCAGCGCCAAAGACUGGUGAUUAUUUCUGAUUG 3497  
Qy 3485 CAGTTCTATTACGATGTATGCTCCCTGGTAAATAGTACTATATCTTAAACAAGTATGCT 3544  
Db 3498 CAGUUUAUAUGAUAAGUGUCUCCAGGCAACAGCACCAUGAUAUUUUUGAUGUCU 3557  
Qy 3545 GTTACCATGAGGTACGTGATAATAGTCTTAATGTGAAGGATTTGTTCTTGAATTTTCC 3604  
Db 3558 GUUACCAUGAGUGUACUGACAUAUUAUGAUAUGUCAAAGUAUGCAUUAUGAUAUGUCU 3617  
Qy 3605 AAAAGTATTCGATGCGCAAGAGAGGTGAACCATGTCTAGAGCCAGTTTTGGTACCOCG 3664  
Db 3618 AAGUCUGUGUGCGCCUUAAGGAUCAAUAUCAAACCAUAUAUUAUUGUACGAAACGCGC 3677  
Qy 3665 GCGGAACCGCAAGGGCTGCAGGACTACTCABAATCTGGTTGCATGTATTAAGAAT 3724  
Db 3678 GCAGAAUUGCCACCGCAGACUGGACUAUUGGAAAUUAUGGCGGAUUAUAAAGGAAC 3737  
Qy 3725 TTCAACGCAACAGACCTGACGCGGACGATTGACATTTGAGAGACCGCATCTGTTGTAGTA 3784  
Db 3738 UUUAAACGACCGGAGUGUGUGCAUUAUGUAUUAUGAUAUUAUGCAUUAUUAUGUUA 3797  
Qy 3785 GATAAGTTTTTGTAGTATTTTAAATAAAGAAATAACAAAAAATAATTTGCTGGA 3844  
Db 3798 GAUAAGUUUUUGAUAUUAUUUCUUAAGAAAGAAACCAAAUAAAUUAUUAUUGUCU 3857  
Qy 3845 GTGATGACGAAGGATTCATGATGAGATGTTGTTGAAACACAGAAAGAGTACTATTGGAC 3904  
Db 3858 UUGUUCAGUAGAGAGUCUCUCAAUAGAUUGUUAAGAAAGCAGGAAACAGUAACAUAGC 3917  
Qy 3905 GACTTGGCTACTACATTTTACAGATCTGCGGCGCATCTGATCAGTCAAGCAACATGATC 3964  
Db 3918 CAGCUGCGAUUUUGAUUUUGUAGAUUUGCCAGCAUUAUGUAUUAUGACACACAUAGU 3977  
Qy 3965 AAGGCTCAACCAAAACAGAAATTTGGACCTTTCAATTCAGAAATGAATCCCTGCTCTGCA 4024  
Db 3978 AAAGCACAACCAAGCAAAAUUUGGACACUUAUCAAUCAAACGGAGUACCCCGCUUGCAG 4037  
Qy 4025 ACAATTTCTACATTCGAGACGATCAACGGTATTTTGGCGGTTTCT---CAGAGCTT 4081  
Db 4038 ACGAUUGUGUACCAUUAUCAAAGAAUCAAUUAUUAUGCCCGUUGUUAUGAGUCUG 4097  
Qy 4082 ACAAGTTTGTCTCGAGGCAATTTGATCTAAGAAATTTCTTTCTTTCTTAGGAAACT 4141  
Db 4098 ACUAGGCAUUAUUGGACAGUUGUUAUUGGAGCAGAUUUUUUUUUUUAUUAUUAUUA 4157  
Qy 4142 CCAGAACAGATTCAGAAATTTTCTCGGATCTCGACTCGCACTGCGCTTCTTATGATGTGTTA 4201  
Db 4158 CCAGCGCAGAUUGAGGAUUUUCUUGGAGAUUUCGACAGUUAUGUGCCGAGUUGUCUUG 4217  
Qy 4202 GAATCTGATATTTCTAAGTATGATAAGTCAAGAACAGATTTCAATTTGCTGTAGAGTAT 4261  
Db 4218 GAGCUGAUAUUAUCAAUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 4277  
Qy 4262 GAAATATGAAAGATTTGGGTCTCAATGAGTTTGGCCGAGGTGTGAAACAAAGGCGAC 4321  
Db 4278 GAGAUUUGCGAAGUUGGUGUUUUAAGACUUUUUGGAGAAUUAUUAUUAUUAUUAUUA 4337  
Qy 4322 AGGAAACCACTTTGAAGGATTTACATTTGCTGGAATCAAGACATGCTGTGTGATCAAGG 4381  
Db 4338 AGAAAGACCCCUCAAGGAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 4397  
Qy 4382 AAAAGCGGTGATGTACTATTTTCATCGGCAATCTGTTATAATAGCAGCTTGTGCTGGT 4441  
Db 4398 AAGACGCGGACGACGACGUGUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 4457  
Qy 4442 TCAATGTTTACGATGGAAGGTCATAAAGGTCCTTTTGTGGAGAGGATTCGCTTTTG 4501  
Db 4458 UCGAUGUCUCCGAGGAGAAAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAU 4517



Qy	740	GTATGTTATGCACGCTTCCATTTTGGCAGAAAGCAATTATTACTAGACACGAGCGAGGTTACG	7399
Db	738	ACGUGCUAGCGCCGCUUUCACUCUCUCUGAAGAACCCUGCUCUUGAAGAUUCAUACGUCAAU	797
Qy	800	CTTTAAATGAATAATGCGCGCAACTTTTCAAAAGAGAGAGGTGATGATGCTTCTTTTCTTTGCT	859
Db	798	UUGGACGAATAACACGGUGUUUUUUCGCGCAUGGAGACAAAGUUGACCUUUUUCUUUUGCA	857
Qy	860	GATGAAAGTACTTTTAAATTATAGTGCATAATAACAAAATATCTTGCATTATGTAGTTAAA	919
Db	858	UCAGAGAGUACUCUUAUAUUGUCAUGUUAUUCUAAUAUUCUUAAGUAUGUGUGCAAA	917
Qy	920	TCTTACTTTCCTGCTCTAGTAGAATAGTTCATTCTTTAAAGAAATTTTGTACTACAGGTT	979
Db	918	ACTUUAUUCCGGCGCUCUUAUAGAGAGGUUUAUCAUGAAGAGGUGUUUAGUCACAGAGU	977
Qy	980	AATACTTGGTTTGTAAATTTTACAAAGTAGATACCTATATCTGTACAAAGAGTGTGTAGA	1039
Db	978	AUAUCUGGUUUUGAAGUUUUCUAGAAUAGAUACUUUUCUUUUUGAACAAAGGUGUGGCC	1037
Qy	1040	CAAGTAGGCTGTGATAGTGCATGAGTCTTATGAGGCGATGGAAGACGCTTTGTCTTACAAG	1099
Db	1038	CAUAAAGUGUAUAUGAGUGACGAGUUAUAUCUGCAUUGGAAGGACGCAUGGCAUUAACAA	1097
Qy	1100	AAAACCTTGCCCATGTTTCAACACTGTAAAGAGCAATCTTTTAGACACACGGCTTCGGTTAAC	1159
Db	1098	AAGACUCUUGCAUUGGCAACAGCGAGAGAAUUCUCUCUGAGGAUUAUCAUCAGACUCAA	1157
Qy	1160	TTTTGGTTCCTTAAGATGAAGGACATGGTATAGTACCGCTGTTTTCAGGGGTTCTATTACC	1219
Db	1158	UACUGGUUUUCCAAAUGAGGGGAUUGGUCACUGCAUUAUUCGACAUUUUCUUUGGAG	1217
Qy	1220	AGCAAAAGATGACAAAGAGTGAGGCTCATTTGTTTAAATCGTGACTTTCGTTTACACAGTCTT	1279
Db	1218	ACTUAGUAAGGACGCGCAAGGAAGUUCUAGUUCUCCAAGGAUUCUGUGUUUACAGUCUU	1277
Qy	1280	AATCATATCAGAAATATCAAGCCAAAGCGTTTAACTTTACAGAAACGTATATCTTTCGTG	1339
Db	1278	AACCAUAUCGAACAUAACAGGCGGAAGCUCUUAUAUCGAAAUUUUUUGUCUUUGUC	1337
Qy	1340	GAGTCTATAAGATCCCGCTGATAATCAATGGTGTACTGCTAGGCTCTGAATCGGATGTA	1399
Db	1338	GAUUCGAUUCGACGAGGGUAUCAUUAACGGUGUGACAGCAGGAGGCCGAUUGGGAGUG	1397
Qy	1400	GATAAAGCAATCTTCAACCCCTGTCAAAGACTTTCTTCTTGACAGACTAAGCTGGCTCGC	1459
Db	1398	GACAAAUUUUGUUAACAUCUUGUCAUGACGUAUUAUCUGCAUAUAAGCUUUGCGUU	1457
Qy	1460	CTTCAAGACGATATAGTAATGGAAGAGTTTCGGTGTCTGGATAAGACCACTTCTGAACATT	1519
Db	1458	CUAAAGGAUGACUUAUGAUUAGCAAGUUAUCUCUGUUCGUAACAAACGGUGGCCAGCAU	1517
Qy	1520	ATTTGGGATGAGGTGGGCAAAATTTTTTGGAAACGTTTTCCCACTATCAAGAGAGATTG	1579
Db	1518	GUGUGGAUGAGAUUUCGUGCGCUUUGGGAACGCAUUCUCCUCCGUGAAGAGAGGCGUC	1577
Qy	1580	GTGACGAGGAAAATCTCTGGATGTAAGTGAATGCTCTGAAGATCAAGATCCACAGATCTG	1639
Db	1578	UUGAACAGGAAAACUUAUCAGAGUGGACGGCGACCAUUAAGAUACAGGGUCCUGAUCUA	1637
Qy	1640	TATGTCACATGGAAGACAGGTTTCGTAGCTGAAATACACCAAGCTCTGAGGAGTTACCGCAT	1699
Db	1638	UAUGUGACCUUCCACGACAGAUUAGUCUGAGUACAAGGCCUCUCUGGAGCAUGGCCUGCG	1697
Qy	1700	CTAGATATCAAGAAAGGACTTAGAAGAGCTGAGCAAAATGTACAGCGGTTATCAGAATTA	1759
Db	1698	CUUGACAUUAGGAAGAAGAUGGAAGAAAACGGAAGUAGUUAUAUGCAUUUCAGAGUUA	1757
Qy	1760	TCTATCCTTAAGGGTCTGATAATTTTCGATATCGCGAAGTTTCAAGACATGTCCAGGCT	1819
Db	1758	UCGGUGUUAAGGGAGUCUUGCAAAUUCUGUUGAUUUUUUUUCCCAAGUUGGCCAAUCU	1817
Qy	1820	TTAGATGTTAGTCTCTGATGTGGCAGCACGAGTAAATCGTTGCAGCTGGCCGAGAAATAGAAGC	1879













3498 CAGUUUUAUUAUUAAGUGUCUCCAGGCAACAGCACCAUGAUAUUAUUAUGUCU 3557  
3545 GTTACCATGAGCTTACGTGATATAGTCTTAATGTGAAGATTGTCTTGATTTTCC 3604  
3558 GUUACCAUGAGGUGACUGACAUUUAUUAUGUUAUUAUGUUAUUAUGUUAUUAUGU 3617  
3605 AAAAGTATTTCCGATGCCAAGAGGTGAACACCATGTCTAGAGCCAGTTTGTGTCACCGG 3664  
3618 AAGUCUGUUGUCGCGCAGAGAUCAUAUCAAACCAUUAUUAUUAUUAUUAUUAUUAU 3677  
3665 GCGGAACCGCCAAAGGCTGACGACTCTCGAAATCTGTGGTTCGAATGATTAAGAAAT 3724  
3678 GCAGAAUUGCCGCGCAGACUGGACUAUUGGAAAUUAUUAUGGCGCAUUAUUAUUAU 3737  
3725 TTCAAGCAGCAGACTGACGCGGACGATTGACATTTGAGACACCGCATCTGTCTGTGTA 3784  
3738 UUUUACGACCGCGAGUUGUCUGCAUCAUUGAUUUUUAUUAUUAUUAUUAUUAUUAU 3797  
3785 GATAAGTTTTTGTATGCTATTTTATTAAGAAAGAAATAACAAAAATATTTCTGGA 3844  
3798 GAUUAUUUUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAU 3857  
3845 GTGATGACGAAGGATTCATGATGATGATGATGATGATGATGATGATGATGATGATG 3904  
3858 UUGUUCAGUAGAGAGUCUCUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAU 3917  
3905 GACTTGGCTAACTACAAATTTTACAGATCTGCGGCGCATCGATCAGTACAGCACATGATC 3964  
3918 CAGCUGCGAGAUUUUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAU 3977  
3965 AAGGCTCAACCAAAACAGAAATTTGACCTTTCAATTTCAAGATGAATCCCTGCTCGCA 4024  
3978 AAGCACAAACCAAGCAAAAUUUGACACUAUUAUUAUUAUUAUUAUUAUUAUUAU 4037  
4025 ACAATGTCTACATTCGAAGCAGATCAACGGTATTTTGGCGGTTTCT---CAGAGTT 4081  
4038 ACGAUUGUGUACCAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAU 4097  
4082 ACAAGTTGCTGCTCGAGCAATTTGATCTTAAGAGTTTCTTTTCTTACTAGGAAACT 4141  
4098 ACUAGCAUAUUAUGACAGUGUUAUUAUGACAGCAUAUUAUUAUUAUUAUUAUUAU 4157  
4142 CCAGAACAGATTCAGAAATTTTCTCGGATCTCGACTCGACGTTTCTTATGGATGTGA 4201  
4158 CCAGCGCAGAUUUGAGAUUUUUCGAGAUUUCGAGAUUUCGAGAUUUCGAGAUUUCG 4217  
4202 GAATGGAATTTCTAAGTATGATAGTCAAGACGAGTTTCTTATGCTGTAGATAT 4261  
4218 GAGCUGGAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAU 4277  
4262 GAAATATGAAAGATTTGGTCTCAATGATTTTGGCGGAAGTGTGGAACAAGGCA 4321  
4278 GAGUUCUGCGAGAUUUGGUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAU 4337  
4322 AGGAAACAACTTTGAAGATTACATGCTGGAATCAAGACATGCTGTGGTATCAAAG 4381  
4338 AGAAGAACCAACCUCAAGAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 4397  
4382 AAAAGGCTGATGTGACTACTTTTCATCGCAATCTGTTATATAGCAGCTTGTGGT 4441  
4398 AAGAGCGGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 4457  
4442 TCAATGTTACCATGAAAGGTCTAAGAGTGTCTTTTGTGGAGCAGTTCCGTTTGTG 4501  
4458 UGAUGUUCUGAGAGAGAAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 4517  
4502 TATTTTCCAAAGGTTTGGATTTCCCTGACATTTCACTGATGCTTAATCTCATGTGAT 4561  
4518 UACUUUCCAAAGGUGUGAGUUCGAGUUGCGAUAUUAUUAUUAUUAUUAUUAUUAU 4577  
4562 TTTGAGCCAACTGTATAGAAGAGGTACGTTACTTTTGTGTAGATACATCATAC 4621  
4578 UTUUGAGCAAAACUGUUUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 4637

4622 CATGATAGGCGCAATAGTGTATATGATCTTTGAAAGTTGATCTCCAAACTTTGGGCA 4681  
4638 CACGACAGAGAUCAUUGUGUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAU 4697  
4682 AAACATATCAAGGATTTATGATCACTTAAAGAGTTAAGGTTGCTTTTGTGCGATTTGCT 4741  
4698 AAACACAUAAGGAGUUGGAAACAUAUGAGAGUUAUUAUUAUUAUUAUUAUUAUUAU 4757  
4742 TGTTCGCTCGGAACTGCTGTAGGCTTTCCGAGCTGAACGAGCTATCAAGAGGTT 4801  
4758 GUUUCGUUGAACAAUUGUGCGUA---UUAACACAGUUGGACGCGUGUAUUGGAGGU 4814  
4802 CATAAAAACCGGATTTGATGTTTCTTTTAAATTTGTTTAAATAAATTTTGTGTGAT 4861  
4815 CAUAAGACCGCCUCCAGGUGUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAU 4874  
4862 AAATTTTATTTAGAACTTTTAAATGCTGTTAGTCTCAGAGATAGTCTCAAAAT 4921  
4875 AAAGUUCUUUUAAGUUGUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAU 4934  
4922 TAGCGAGTTTATGATCTTTTCAACAGATGAGATCTTCCGCTATTCATGACTAAGT 4981  
4935 CAUGAGUUUAUUGACCGACAAUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAU 4994  
4982 CAAGAGTCTTAAATATCTGACTGTGGAACAAGATTATGCTGTTTAAAGATGATGCTTTT 5041  
4995 AAAGAGUUAUUGUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 5054  
5042 TGATGATGATTTTAAAGGTTTAAATGTTTAAAGAAAGGTTATGTTGCTTGTAGTGA 5101  
5055 AGAGGUAACCUUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAU 5114  
5102 TTTGTGATGCTGCGGAGTGAATCTCCGATAACTCCGCTGCTGCTGCTGCTGCTGCTG 5161  
5115 UUGUUCUGUACGCGGAGUGGAAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 5174  
5162 TATTGTAGATAAGAGATGAAGAGGATGAAGAGCAACGCTGGTGGTATCACGCCCC 5221  
5175 UCUGUGGACAAAGGAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 5234  
5222 TGCTTGGCAAAAGAAATTTTCTTTTAACTTAATCCCTAATTTTCAATTAACATCCGAG 5281  
5235 AGUCGAAAGAAAGAAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 5294  
5282 TGCTGAGAGACACCGTGCAGTGTAGTGAATCAATCAAGAGTGGCTATCGAAGAGG 5341  
5295 CCGAUGAAACGUCUGGCAAGUUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 5354  
5342 ATACTGCTCTTTATCTTTGGAGTTCTGTTTCAATTTTGTGTAGTACATAAATAATGTAAG 5401  
5355 UUCUGUGCGCUUUCUGGAGUUUGUGUGUGUGUAUUAUUAUUAUUAUUAUUAUUAU 5414  
5402 AAAAGTTTGGAGGAAACGATTTTGTAGTGTGACAGACGCTCGCAATTTGAATCACTGA 5461  
5415 AUUAGUUUGAGAGAGAAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 5474  
5462 AAAAGTTTGTGAGGATTTCTGATGAAGTACCAATGCTGTGAATCGAAAAAGTTTCC 5521  
5475 AGAAGUUCUGAUGAGUUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAU 5534  
5522 GG-----AAAAAAGAAATGTTAGTAAATTAATGTAATAAT----- 5561  
5535 AUUCGAACCGGAAAGAGAGUUGUGCGCAAGGAGGAAAUUAUUAUUAUUAUUAUUA 5594  
5562 -----AAGAAATAATACAGTGTGAAGAGGTTTAAATTTAGGAATTTGAGGA 5614  
5595 AGUGCCGAACAGAAUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 5654  
5615 TAATGTAAGTACGAGT-----CTATCCGCTCATCGAGTACGCTTTTAAATCAAT 5665  
5655 UAAUUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 5711

```
QY 5666 ATGCTTATACATCACTCTCCGAGCCAAATTTGTTTACTTATCTTCCTCGCTTACCGAGAT 5725
Db 5712 AUGUCUACAGUAUCACUACUCAGUCUGUGUUCUACUACGAGCGGCGCCGAC 5771
QY 5726 CCGTGTGAGCTGATCAATCTGTGTACAAATGCTGGTAAACAGTCTTCAAAACGCAACA 5785
Db 5772 CCNAUAGAGUUAUUAUUUUGUACUAGUCCUAGGAGGAGUUAUUAUUAUUAUUAUUAUUA 5831
QY 5786 GCTAGGACAAACAGTCCAAACAGCAATTTGGGATGCTGTGAAACCTGTGCTAGTATGACA 5845
Db 5832 GCUCGAAACUGUGUUCAAAGACAAUUCAGUGAGGUGUGGAAACCUUCACCAACAGUAACU 5891
QY 5846 GTGAGATTCCTGATCGGATTCCTATGTGTATAGATATTAATTCGACGCTTGATCGTTG 5905
Db 5892 GUUAGUGUCCUGACAGUACUUAAGGUGUACAGUACAAUUGCGGUUAUUAAGACCCGUA 5951
QY 5906 ATCAGCGGCTTATTAATAGCTTTGATCTAGAAATAGAAATAGAGGTTGATATCA 5965
Db 5952 GUACAGACUUGUAGGUGUACUACUACUAGAAUAGAAUUAUUAUUAUUAUUAUUAUUAUUA 6011
QY 5966 CCGGACCGAATATCTAGTAATCTTAACCGGACTCAGAGGTTAGACGATGCTACTGTA 6025
Db 6012 GCGAACCCACGACUGCCGAAACGUUAGUAGUACUUGUAGUAGUAGUAGUAGUAGUAGUAG 6071
QY 6026 GCTATAAGGGCTTCAATCAATTAATTTGGCTAATGAACTGTTCTGGAACTGGCATGTT 6085
Db 6072 GCUAAGAGGCGCGGAUAAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 6131
QY 6086 AATCAAGCAGGCTTTGAGACTGCTAGTGTGACTGTCTGACCAACCTCCGCTACTTAG 6145
Db 6132 AAUCGAGCUUUCGAGAGCUUCUGUGUUGUUGUUGUUGUUGUUGUUGUUGUUGUUGUUG 6191
QY 6146 CTATT 6150
Db 6192 CAUUU 6196

RESULT 15
US-09-259-741-5
Sequence 5, Application US/09259741
Patent No. 603895
GENERAL INFORMATION:
APPLICANT: GARTER, STEPHEN
APPLICANT: HOLTZ, R. BARRY
APPLICANT: MCCULLOCH, MICHAEL
APPLICANT: TURPEN, THOMAS
TITLE OF INVENTION: A PROCESS FOR ISOLATING AND
TITLE OF INVENTION: PURIFYING VIRUSES SOLUBLE PROTEINS AND PEPTIDES FROM PLANT
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESS: Howrey & Simon
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/259,741
FILING DATE: February 25, 1999
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/037,751
FILING DATE: March 10, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Halluin, Albert P
REGISTRATION NUMBER: 25,277
REFERENCE/DOCKET NUMBER: 00801.0140.US01
```

```
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-463-8100
; TELEFAX: 650-463-8400
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6446 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: Genomic RNA
US-09-259-741-5

Query Match 38.5%; Score 2449.8; DB 3; Length 6446;
Best Local Similarity 44.8%; Pred. No. 0;
Matches 2771; Conservative 1146; Mismatches 2212; Indels 60; Gaps 8;

QY 20 ACAACAACAATTAATAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 79
Db 18 ACCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 77
QY 80 ATACAATCTATAATTTAGCAACGCCCTTTCTTGAAGCGTGAGTGTGTAATAAACAACAACA 139
Db 78 ACACAGACAGCUACCAACAACAGCUUUGUGACACUUGCCGAGGAAACAACUCCUGGUC 137
QY 140 AATGACCTTGCAAGAGCGCATGTACGATACGGCCGTGGAAGAATTTAAACGCCCGCAC 199
Db 138 AAUGAUCUAGCAAGCGUGUCUUUACGACACAGCGUGUAGAGUUUAACGUCUGGAC 197
QY 200 CGTAGACCAAGGTCACATTTTCCAAACTATTAGCAAGAGCAACGCTTCTAGTCTCC 259
Db 198 GCGAGGCCAAGGUGAACUUUUAAGAAUAAUAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 257
QY 260 AACGCTACCCGGAGTTCAGATTACCTTTTATATATCTCAAAATGCGGTACACAGTTTG 319
Db 258 CGGGGUAUCCAGAAUCCAAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 317
QY 320 GCTGAGGTTTGAGACATTTAGAAATTTGAAATATCTGATGCTACAAGTTCCCTATGATCG 379
Db 318 GCAGGUGAUGGUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 377
QY 380 CCACATATGATAGTGGGAACTTTGACACACATTTGTTCAAGGCGAGGATTTACGTG 439
Db 378 UUGACUUAUAGAAUAGGCGGAAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 437
QY 440 CATTGCTGTATGCGCAATCTCGACATACGAGATATATGAGGCGACGAGGAGGAGGAGGAGGAG 499
Db 438 CACUGCUGAUGCCCAACCCUGACGUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 497
QY 500 TCAATTGAGATGATTTGTCAGATTTGTCGTTCTTAAACAAGGTAATTTCTGAGTTTCAA 559
Db 498 AGUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 557
QY 560 AGGAGGCTTTTAACAGGTATGAGAGGCTCCCAACGAAGTCTGCTCTCTAAAATTTT 619
Db 558 AAGGAAGCAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 617
QY 620 CAGGATTTGCAATATATCCCGCAGAGATAGTGTAGAGATACGCTGTTGCTCTGCAC 679
Db 618 CAGACAAUGCGACUACAGCCGCAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 677
QY 680 AGTTGTATGATATTTCTGTGATGAGTTTGAGGCTGCTGTTAATATCTAAGATATATAT 739
Db 678 AGCAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 737
QY 740 GTATGTTATGAGCTTCCATTTTGGCAGAACATTTACTAGACCAACGAGGATTTAG 799
Db 738 ACUGGCUAUGCCGCUUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 797
QY 800 CTTAATGAATAGCGCAACTTTCAAAAGAGAGAGTGTATGATGTTCTTTTCTTTCTTTCTT 859
Db 798 UUGGACGAAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 857
```



Db 3018 UUGUUAUGAGCAUUCUUCGUUUAUAAAGUUCUUCGAGGAGUUGUGCGGAGCC 3077  
Qy 3065 GGAGTACTAAACAGTGTTCACAAACCACTAAAGAGGAAATTTGAATCTTCACTCAGGCT 3124  
Db 3078 GCCGUGAUAUCCGUAUCUCAAACCCUUGCAUGGCAAGUCCUGAUCUUAUCCAAUUG 3137  
Qy 3125 GATAAATTTAGTTAGAGGAGAGGGCTATAGATGTGAACACCGTTCATGAGATCCAA 3184  
Db 3138 GAUAAAGAGCUCUGUCUUUCAAGAGGUAUUCAGAUUGUUCACUGUGCAUGAAGUGCAA 3197  
Qy 3185 GGAGAAACCTTTGAAGATGTGCTGTGAGTGAAGGCACTTCCACTGACTCTGAT 3244  
Db 3198 GCGGAGCAUACUGAUGUUCACUAGUAGGUUAUACCCUACACAGUUCUUAU 3257  
Qy 3245 TCCAAGTCTCCCGCATGTTCTAGTCTGTAGTACAGACACAAAGAGCTTCAATAT 3304  
Db 3258 GCAGGAGACGCCCAUGUUGUGUCAUUGCAUAGGCAACACUUGUCUUAAGUAC 3317  
Qy 3305 TACACCGTAGTTAGATCCTTTAGTACAGATATTTAGTGTGTTCTTTTAAGTCC 3364  
Db 3318 UACACUGUUGUUAUGGAUCCUUAUAGUUAUUAUAGAGAUUCUAGAGAAACUUAUGCUG 3377  
Qy 3365 TTCTCTTTAGAAATGTAATGTAGACAGTGTAGTACAGATACAGATGATGCA 3424  
Db 3378 UACUUGUUAUUAUAGUUAAGUUGCAUGAGGACACAAUAGCAUUAUACAUUGACUG 3437  
Qy 3425 GTGTTCAAAGGTCTAATCTCTTTGTGGCAACACCTAAATCAGGAGACTTCCAGATCTA 3484  
Db 3438 GUGUUAAGGUUCCAAUUCUUGUUGCGCGCAAGACUUGUGUAUUAUUCUGAUAUG 3497  
Qy 3485 CAGTCTATTAGATGATGCTCCCTGGTAAATAGTACTATCTTAAACAGATGATGCT 3544  
Db 3498 CAGUUAUUAUAGUUAAGUUGUCUCCAGGCAACAGCACCAUGAUGAUAUUAUUGAUGCU 3557  
Qy 3545 GTTACCATGAGTTACGTGATAATAGTCTTAAATGGAAGGATGTGTTCTTGAATTTTCC 3604  
Db 3558 GUUACCAUGAGUUGAUGAUGAUAUUAUGUUAUUAUGUUAUUAUGAUAUGU 3617  
Qy 3605 AAAAGTATTCGATGCCAAGAGGTGAACCATGTCTAGAGCCAGTTTTCGTCACGCG 3664  
Db 3618 AAGUCUGUUGCGCCUAGGAUCAAUCAAACCAUUAUUAUUGUUAUUAUGUUAUUAUGU 3677  
Qy 3665 GCGGAACCGCCNAGGGCTGAGGACTACTCGAAATCTGGTTGCAATGATTAAGAAT 3724  
Db 3678 GCAGAAUUGCCACGCGCAGACUGGACUUAUUGGAAAUUAUUGGCGGAUUAUUAAGGAAC 3737  
Qy 3725 TTCAACGACACGACTGACGGGAGGATGACATTTGAGAGCACGCGATCTGTTGTAGTA 3784  
Db 3738 UUAACGCAACCGAGUUGUUGGCAUUAUGAUAUUAUGAUAUUAUGAUAUUAUGAUAUGA 3797  
Qy 3785 GATAAATTTTGTAGTCTTTTATTAATAAAGAAATAACACAAATAATTTGCTGGA 3844  
Db 3798 GAUAAAGUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAU 3857  
Qy 3845 GTGATGACGAGGATTAATGATGATGTTGGAAACACAGAAAGATGACTATTGGAC 3904  
Db 3858 UUGUUCAGUAGAGUUCUCAUUAUUAUGUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 3917  
Qy 3905 GACTTGGCTTAACATACTTTTACAGATCTGCGGCGCATCTGATGATGATGATGATGATG 3964  
Db 3918 CAGCUGGAGAUUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 3977  
Qy 3965 AAGGCTCAACAAACAGAAATTTGACCTTTCAATTTAGAAATGAATACCTGCTCTGCAA 4024  
Db 3978 AAAGCAACCCCAAGCAAAAUUGGACACUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 4037  
Qy 4025 ACAATGCTTACATTCGAGCAGATCAACGGTATTTTGGCGGTTTCT---CAGAGCTT 4081  
Db 4038 ACGAUUGUGUACCAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 4097  
Qy 4082 ACAAGTGTCTGCTCGAGCAATTCATCTAGAGATTTCTTTTCTTACTAGGAAACT 4141  
Db 4098 ACTUUGCAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 4157

Qy 4142 CCAGAACAGATTCAAGAAATTTTCTCGATCTCGACTCGACAGTTCCTATGATGTGTTA 4201  
Db 4158 CCAGCGCAGAUUGAGGAUUCUUCGGAUUCUOCAGCAGUCAUGCGCGGAUGUGUCUG 4217  
Qy 4202 GAATCGATATTTCTAAGTATGATAAGTCAAGAAACGAGTTTCTATGCTGTAGAGTAT 4261  
Db 4218 GAGCUGAUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 4277  
Qy 4262 GAAATATGAAAGATTGGGTCTCAATGAGTTTGTGGCCGAGTGTGGAACAAGGGGAC 4321  
Db 4278 GAGAUUCGCGCAUUGGUGUUAAGAAGUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAU 4337  
Qy 4322 AGGAAACAATTTGAAGGATTACATTTCTGGAATCAAGCATCTCTGTGATATCAAGG 4381  
Db 4338 AGAAAGACCAACCCUACAGGAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 4397  
Qy 4382 AAAAGCGGTGATGTACTTTCTATCGGCAATCTGTATAATAGCAGCTTGTCTGGGT 4441  
Db 4398 AAGAGCGGGACGUCACGACUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 4457  
Qy 4442 TCAATGTTACGATGGAAGGTCTATAAAGGTCTTTTGTGGAGAGATTCGTTTGTG 4501  
Db 4458 UCGAUGCUCUCCGAGGAGAAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 4517  
Qy 4502 TATTTTCAAAAGGTTGGATTTCCCTGACATTCAGTCTGCTATCTCTCATGTGAT 4561  
Db 4518 UACUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 4577  
Qy 4562 TTTGAGGCCAACTGTATAGAAAGGTTACGGTTTCTTTGTGTAGATACATCATACAC 4621  
Db 4578 UUGAGCAAAACUUGUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 4637  
Qy 4622 CATGATAAGGAGCAATAGTGTATATGATCCTTTGAAAGTTGATCTCAAACTTTGGGCA 4681  
Db 4638 CACGACAGAGCAUUGUGUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 4697  
Qy 4682 AAAATATCAAGGATTTGATCACTTAGAAGATTAGGGTGTCTTTGTGCGATGTTGCT 4741  
Db 4698 AAACACAUCAAGGAUUGGAAACACUUGAGGAGUUCAGAAAGGUCUUAUUAUUAUUA 4757  
Qy 4742 TGTTTCGTCGGAACCTGTGTCTTAGGTTTCCGACAGCTGAAACGACGCTATCAAGAGGTT 4801  
Db 4758 GUUUGUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 4814  
Qy 4802 CATAAAACCGCATGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 4861  
Db 4815 CAUAAAGCGCCGCCUCCAGGUGUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 4874  
Qy 4862 AAATTTTATTAGAACTTTGTTTAAATGGCTGTAGTCTCAGAGATCTGTCAAAAT 4921  
Db 4875 AAAGUUCUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAU 4934  
Qy 4922 TAGCGAGTTCAATGATCTTTGAAACAGATGAGATCTTCCGGCATTCATGACTAAGT 4981  
Db 4935 CAUAGUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 4994  
Qy 4982 CAAGAGTGTAGAAATATCGACTGTGGAACAAGTATGCTGTGTAAGATGATGCTTTC 5041  
Db 4995 AAAGAGUUAUUGUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 5054  
Qy 5042 TGATGTAGATTATTAAGGTTGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAG 5101  
Db 5055 AGAGUGAAACCTUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 5114  
Qy 5102 TTTGCTAGTGTCTGGGAGTGAATCTCCCGATAACTCCGCTGTGTGTGTGTGTGTGTGTGT 5161  
Db 5115 UUGUGUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAU 5174  
Qy 5162 TATTGTAGTAAAGAAATGAAAGAGGTAAAGAAAGCAACGCTGGGTGCGTATACGCGCC 5221  
Db 5175 UCUGUGGACAAAGGAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 5234





Result No.	Query			DB	ID	Description
	Score	Match	Length			
C 1	303.4	4.7	702	4	BM068217	KS08018P1
C 2	299.8	4.7	681	4	BM058406	KS08022B0
C 3	299.6	4.7	700	4	BM067779	KS08011B0
C 4	299.6	4.7	717	4	BM068367	KS08022E0
C 5	297.6	4.7	668	4	BM068623	KS08025D0
C 6	296.6	4.7	663	4	BM068294	KS08020D0
C 7	296.2	4.7	663	4	BM068316	KS08020P1
C 8	295.8	4.7	701	4	BM067993	KS08015C0
C 9	295.6	4.7	665	4	BM067343	KS08003C0
C 10	293.8	4.6	660	4	BM067946	KS08014D0
C 11	293.6	4.6	693	4	BM068203	KS08018E0
C 12	292.8	4.6	656	4	BM068536	KS08025F0
C 13	292.6	4.6	658	4	BM068286	KS08020C0
C 14	292.6	4.6	677	4	BM067601	KS08007H1
C 15	292.4	4.6	679	4	BM067465	KS08005D1
C 16	291.2	4.6	655	4	BM067937	KS08014C1
C 17	290.2	4.6	654	4	BM067953	KS08014E0
C 18	290.2	4.6	694	4	BM068482	KS08024F0
C 19	289.8	4.6	652	4	BM067532	KS08006H0
C 20	289.8	4.6	665	4	BM068348	KS08022C0
C 21	287.4	4.5	658	4	BM067756	KS08010E0
C 22	287.2	4.5	651	4	BM067815	KS08011H0
C 23	286.2	4.5	667	4	BM068012	KS08015E1
C 24	286	4.5	654	4	BM067516	KS08006E0

QY 3206 TCCTGGTCAGATTGACGCAACTCCACTGACTCTGATTTCCAAGTCTTCCCGCATGTT 3265  
Db 619 TCATTAGTGAGACTGACCGCTACACCGTGGGAATAATTTCAAAGCAGAGTCCGACCTG 560  
QY 3266 CTAGTCGCTCTGACTAGACACAAAGAGCTTCAATAATTACACCGTAGTGTAGATCCT 3325  
Db 559 TTGGTCTCGTGTCTAGGAGATCAAGGTCATCAATAATTACACTGTTGTACTAGATGCA 500  
QY 3326 TTAGTACAGATAATTAGTGTCTTCTTTAAAGCTCTCTTAAAGCTCTCTTTAGAAATGATATG 3385  
Db 499 GTCGTTTCACTTCTAGAGATTGGAGTGTGTAGTAGTCTCTTGTAGATATGACAAA 440  
QY 3386 GTAGAGCAGGAGTAGATAGACAAATTACAGATGAGTGTCTTCAAAGGTCAATATCTC 3445  
Db 439 GTTGATGTCTGACTCAATAGCAATACAGATAGAAATCGGTGTACAAAGGTGTAACTT 380  
QY 3446 TTGTGGCAACACCTTAATTCAGGAGACTTCCAGATCTACAGTCTTATACAGATGATGC 3505  
Db 379 TTCGTGCGAGCCCCGAAACAGAGATGTTCTGACATGCAATATTAATGACAAAGTGT 320  
QY 3506 CTCCTGTGTAATAGTACTATCTTAACAAAGTATGCTGTTPACCATGAGGTTACGTGAT 3565  
Db 319 TTGCGGGGAAACAGTACTATCTTAATGATATGATGCTGTAACTATGCAATACGAGA 260  
QY 3566 AATAGTCTTAATGTGAAGATTGTGTTCTTGAATTTTCCAAAGTATTCOGATGCCAAG 3625  
Db 259 AATAATTTGAATGTCAAGGATTGTGTGATATGTCGAGTGTGCTGCTTCCGAGA 200  
QY 3626 GAGGTGAACACCTGCTAGAGCCAGTTTGGCTACCGGGGAGACCCGAGGCTGCA 3685  
Db 199 GAATCTGAGACAGATTGAACTGTGATCAGGACTGCTGTGCTGAAACCTCGAAACCT 140  
QY 3686 GGACTACTCGAAATCTGGTTGCAATGATTAAGAAATTTCAACGCGACCAAGCTGACG 3745  
Db 139 GGATTTGGAAACTTGGTCGAGATGATCAAGAAATTTCACTCTCCGAAATTAATA 80  
QY 3746 GGGAGATTGACATGAGAGCAGCCGATCTGTTGTAGTAGATAAGTTTTTTGTAGCTAT 3805  
Db 79 GGGGTGCTGATCATCGAAGACACCGCTTCTCTAGTAGTAGATAAGTTTTTTGTAGCTAT 20  
QY 3806 TTTATTAATAAAG 3818  
Db 19 TTTATTAATAAAG 7

RESULT 2  
LOCUS BM068406/c 681 bp mRNA linear EST 11-SEP-2002  
DEFINITION KS08023B09 KS08 Capsicum annuum cDNA, mRNA sequence.  
ACCESSION BM068406  
VERSION BM068406.1 GI:22788526  
KEYWORDS EST.  
SOURCE Capsicum annuum  
ORGANISM Capsicum annuum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
asterids; lamiales; Solanales; Solanaceae; Capsicum.  
1 (bases 1 to 681)  
Lee, S., Kim, S.-Y., Chung, Y.-H., Shin, H.-J., Goh, S.-H., Pai, H.-S.,  
Hur, C.-G. and Choi, D.  
Generation of Expressed Sequence Tags from Hot Pepper (Capsicum  
annuum L.) and Sequence Analysis in Relation to Hypersensitive  
Response Against Pathogen  
Unpublished (2001)  
Contact: Doil Choi  
Genome Research Center and National Center for Genome Information  
Korea Research Institute of Bioscience and Biotechnology  
P.O. Box 115, Yuseong, Taejeon, 305-600, Republic of Korea  
Tel: 82-42-860-4340  
Fax: 82-42-860-4309  
Email: doil@mail.kribb.re.kr  
High quality sequence stop: 681.  
Location/Qualifiers

## FEATURES

source 1. 681  
/organism="Capsicum annuum"  
/mol\_type="mRNA"  
/cultivar="Hang Keun"  
/db\_xref="taxon:4072"  
/tissue\_type="anther"  
/dev\_stage="10 weeks after germination"  
/clone\_lib="KS08"  
/note="Vector: pBluescript SK(-)"  
ORIGIN  
Query Match 4.7%; Score 299.8; DB 4; Length 681;  
Best Local Similarity 65.4%; Pred. No. 2.2e-62;  
Matches 439; Conservative 0; Mismatches 232; Indels 0; Gaps 0;  
QY 3146 AAGCGCTATAAGATGTGAACACCGCTTCATGAGATCCAGGAGAAACCTTTGAAGATGTG 3205  
Db 678 AGGGGTTTACGAAAGATGTCATACTGTTTCATGAGGTGCAAGGGGAAACGTTTGAAGACGTC 619  
QY 3206 TCGCTGGTTCAGATTGACGGCAACTCCACTGACTCTGATTTTCCAAGTCTTCCCGCATGTT 3265  
Db 618 TCATTAGTGAGACTGACGCTTACACCGTGGGAATAATTTCAAAGCAGAGTCCGACCTG 559  
QY 3266 CTAGTCGCTCTGACTAGACACAAAGAGCTTCAATAATTACACCGTAGTGTAGATCCT 3325  
Db 558 TTGCTCTCGTGTCTAGGCATACAAAGTCAATCAATATTAATTAATTAATTAATTAATTA 499  
QY 3326 TTAGTACAGATAATTAGTGTGTTCTTCTTAAAGTCTTCTTCTTAAAGTCTTCTTCTTAA 3385  
Db 498 GTCGTTTTCAGTGTCTAGAGATTGGAGTGTGTAGTAGTACCTGTGTAGATATGACAAA 439  
QY 3386 GTAGAAGCAGGTAGTAGATAGCAATTAACAGATGATGATGATGATGATGATGATGATGATG 3445  
Db 438 GTTGATGTCGACTCAATAGCAATTAACAGATGATGATGATGATGATGATGATGATGATGAT 379  
QY 3446 TTGTGCGAACACCTAAATCAGGAGCTTTCCAGATCTACAGTCTTATTAAGATGATGATG 3505  
Db 378 TTGTCGCGAGCCCGAAACAGGAGATGTTTCTGACATGCAATATTAATTAATTAATTAATTA 319  
QY 3506 CTCCTCGTAAATAGTACTACTTAAACAGTATGATGATGATGATGATGATGATGATGATGAT 3565  
Db 318 TTGCGGGGAAACAGTACTACTTAAAGTATGATGATGATGATGATGATGATGATGATGATGAT 259  
QY 3566 AATAGTCTTAATGTGAAGATTGTGTTCTTGAATTTTCCAAAAGTATTCGATGCCCAAAG 3625  
Db 258 AATAATTTGAATGTCAAGGATTGTGTGATATGATGATGATGATGATGATGATGATGATGAT 199  
QY 3626 GAGGTGAACCATGCTAGAGCCAGTTTTCGCTACCGCGGCGGAAACCGGCAAGGCTGCA 3685  
Db 198 GAATCTGAGACGACATTTGAACCTGTGATCAGGACTGCTGCTGAAACACCTCGAAAAACCT 139  
QY 3686 GGACTACTCGAAAAATCTGGTTGCAATGATTAAGAAATTTCAACGCGACCAAGCTGACG 3745  
Db 138 GGATTTGGAATACTTGGTCGATGATCAAGAAATTTCACTCTCCGAAATTAATA 79  
QY 3746 GGGACGATTGACATGAGAGCAGCAGCTGTTGTGTAGTAGATAAGTTTTTTGTAGTAGCTAT 3805  
Db 78 GGGGTCTGTGACATCGAAGACACCGCTTCTCTAGTAGTAGATAAGTTTTTTGTAGTAGCTAT 19  
QY 3806 TTTATTAATAAAG 3816  
Db 18 TTTATTAATAAAG 8

RESULT 3  
LOCUS BM057779/c 700 bp mRNA linear EST 11-SEP-2002  
DEFINITION KS08011B03 KS08 Capsicum annuum cDNA, mRNA sequence.  
ACCESSION BM057779  
VERSION BM057779.1 GI:22787899  
KEYWORDS EST.  
SOURCE Capsicum annuum  
ORGANISM Capsicum annuum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; lamids; Solanales; Solanaceae; Capsicum.

REFERENCE 1 (bases 1 to 700)

AUTHORS Lee, S.-Y., Kim, S.-Y., Chung, Y.-H., Shin, H.-J., Goh, S.-H., Pai, H.-S., Hur, C.-G. and Choi, D.

TITLE Generation of Expressed Sequence Tags from Hot Pepper (Capsicum annuum L.) and Sequence Analysis in Relation to Hypersensitive Response Against Pathogen

JOURNAL Unpublished (2001)

COMMENT Contact: Doil Choi

Genome Research Center and National Center for Genome Information

Korea Research Institute of Bioscience and Biotechnology

P.O. Box 115, Yusong, Taejeon, 305-600, Republic of Korea

Tel: 82-42-860-4340

Fax: 82-42-860-4309

Email: doil@mail.kribb.re.kr

High quality sequence stop: 700.

FEATURES

source

1..700

/organism="Capsicum annuum"

/mol\_type="mRNA"

/cultivar="Hang Keun"

/db\_xref="taxon:4072"

/tissue\_type="anther"

/dev\_stage="10 weeks after germination"

/clone\_lib="KS08"

/note="Vector: pBluescript SK(-)"

ORIGIN

Query Match 4.7%; Score 299.6; DB 4; Length 700;  
Best Local Similarity 64.6%; Pred. No. 2.5e-62;  
Matches 446; Conservative 0; Mismatches 244; Indels 0; Gaps 0;

QY 3110 ACTTTCACTCAGGCTGATAAATTTGAGTTAGAGGAGAGGGCTATAAGATGTGAACACC 3169  
DB 699 ACATTCATCTCAATCAGCAAGTCATCTGCTCTCAAGGGGTTACAGATGTGATACC 640

QY 3170 GTTCATGAGATCAAGGAGAAACCTTTGAAGATGTGCTCGTGTGAGATTTGACGGCACT 3229  
DB 639 GTTCATGAGGTCAAGGGGAAACGTTTGAAGAGCTCTCATTTAGTGTGAGACTGACGCTACA 580

QY 3230 CCACTGACTCTGATTTCCAGTCTCCCGGATGTCTTGTAGTGTGCTGTGACTAGACACACA 3289  
DB 579 CCGCTGGGAATAATTTCAAGCAGAGTCCGCACCTGTTGTGCTCGTTGTCTAGGCATACA 520

QY 3290 AAGAGCTTCAATATTTACACCGTAGTGTGTAGATCCTTTAGTACAGATTAATAGTATTG 3349  
DB 519 AGGTCAATCAATATTTACACTGTTGTACTAGATGAGTGTTCAGTGTCTTAGAGATTG 460

QY 3350 TCTTCTTTAAGCTCCTCTCTTTTGAATATGTATGTGTAGAGCAGGTAGTAGATAGCAA 3409  
DB 459 GAGTGTGTGAGTGTACCTGTTAGATATGTACAAAGTTGATGTGCTCAATAGCAA 400

QY 3410 TTACAGATGGATGAGTGTTCAGAGTCAATATCTTTTGTGGCAACACCTTAATCAGGA 3469  
DB 399 TTACAGATAGAAATCGGTGTACAAAGGTGTAACTTTTCGTGCGAGCCCCGAAACAGGA 340

QY 3470 GACTTTCAGATCTACAGTCTTATACGATGTATCCTCCCTGTTAATAGTACTACTATT 3529  
DB 339 GATGTTTCTGACATGCAATATTACTATGATCAAGTGTTCGCGGAAACAGTACTACTATT 280

QY 3530 AACAAATATGATGCTGTTTACCATGAGTTTACGTGTATATAGTCTTAAATGGAAGATTGT 3589  
DB 279 AATGATATGATGCTGTAACTATGCAATACGAGAAATAATTTGAAATGCAAGATTGT 220

QY 3590 GTTCTTGATTTTCCAAAGATATTCGATGCCAAAGAGGTGAACACCATGTCTAGAGCCA 3649  
DB 219 GTTGTGATATGTGGAATCGGATCGGTGCTCTTCCGAGAGAAATCTGAGACGACATTGAAACCT 160

QY 3650 GTTTCGTCACCGCGGAAACCGCAAGGCTGACGAGTACTCGAAATCTGTTGCA 3709  
DB 159 GTGATCAGGACTGCTGCTGAAAAACCTCGAAACCTGGATTGTGGAAACCTTGGTCGG 100

QY 3710 ATGATTAAAGAAATTTCAACGCCACGACCTGACGGGAGCATTTGACATTGAGAGCACC 3769  
DB 99 ATGATCAAAAGAAATTTCAACTCTCCCGAATTAATAGGGGTGCTTGACATCGAAGACACC 40

QY 3770 GCATCTGTTGTAGTAGATAAGTCTTTTTCAT 3799  
DB 39 GCTTCTCTAGTAGTAGATAAGTCTTTTTCAT 10

RESULT 4

BM068367/c

LOCUS

BM068367

DEFINITION

BM068367

VERSION

BM068367.1

KEYWORDS

EST.

SOURCE

Capsicum annuum

ORGANISM

Capsicum annuum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

asterids; lamids; Solanales; Solanaceae; Capsicum.

REFERENCE 1 (bases 1 to 717)

AUTHORS Lee, S.-Y., Kim, S.-Y., Chung, Y.-H., Shin, H.-J., Goh, S.-H., Pai, H.-S., Hur, C.-G. and Choi, D.

TITLE Generation of Expressed Sequence Tags from Hot Pepper (Capsicum annuum L.) and Sequence Analysis in Relation to Hypersensitive Response Against Pathogen

JOURNAL Unpublished (2001)

COMMENT Contact: Doil Choi

Genome Research Center and National Center for Genome Information

Korea Research Institute of Bioscience and Biotechnology

P.O. Box 115, Yusong, Taejeon, 305-600, Republic of Korea

Tel: 82-42-860-4340

Fax: 82-42-860-4309

Email: doil@mail.kribb.re.kr

High quality sequence stop: 717.

FEATURES

source

1..717

/organism="Capsicum annuum"

/mol\_type="mRNA"

/cultivar="Hang Keun"

/db\_xref="taxon:4072"

/tissue\_type="anther"

/dev\_stage="10 weeks after germination"

/clone\_lib="KS08"

/note="Vector: pBluescript SK(-)"

ORIGIN

Query Match 4.7%; Score 299.6; DB 4; Length 717;  
Best Local Similarity 64.8%; Pred. No. 2.5e-62;  
Matches 460; Conservative 0; Mismatches 249; Indels 1; Gaps 1;

QY 3108 TAACTTTCACCTCAGGCTGATAAATTTGAGTTAGAGGAGAGGGCTATAAGATGTGAACA 3167  
DB 717 TTACATTCACCAATCAGACAGTCATTTGCTCTCAAGGGGTACGAAGATGTGCATA 658

QY 3168 CCGTTTCATGAGATCCAAAGGAGAAACCTTTGAAGATGTGTGCTGTGCTCAGATTGACGGCAA 3227  
DB 657 CCGTTTCATGAGGTGCNAGGGGAAAGCTTTGAAGAGCTCTCATTTAGTGTGAGACTGACGCTTA 598

QY 3228 TCCCACTGACTCTGATTTCCAAAGCTTCCCGCATGTTCTAGTCTGCTGTGCTAGACACA 3287  
DB 597 CACCGTGGGAATAATTTCAAAGCAGAGTCCGCACCTGTGTGCTCTGTTGTCTAGGCATA 538

QY 3288 CAAAGAGCTTCAATATTTACACCGTAGTGTGTAGATCCCTTTAGTACAGATAATTAGTGATT 3347  
DB 537 CAAAGTCAATCAATATTTACACTGTGTGTACTAGATGCAGTCGTTTTCAGTGTCTAGAGATT 478

QY 3348 TGTCTTCTTTTAAAGCTCTCTCTTTTAAAGAAATGTATATGTGAAGCAGAGTGTGCTCATTAGC 3407  
DB 477 TGGAGTGTGTAGTAGTTACCTGTTTAGATATGTACAAAGTTGATGTGTGCTCATTAGC 418

QY 3408 AATTACAGATGATCGAGTGTTCAAAGG-TCATAATCTCTTTGTGGCAACACCTAAATCA 3466  
 |||||  
 Db 417 AATTACAGATAGATCGGTGTACAAAGGNTGTTAACTTTTCGTCGACGCCCGGAAACA 358  
 |||||  
 QY 3467 GGAGACTTTCAGATCTCAGTCTCTATTACAGTGTATCCCTCCCTGTTAAATAGTACTATA 3526  
 |||||  
 Db 357 GGAGATGTTTCTGACATGCAATATTACTATGACAAAGTGTTCGCGGGAACAGTACTATA 298  
 |||||  
 QY 3527 CTTAAAGATGATGCTGTTACCATGAGGTTTACGTGATAATAGTCTTAAATGTGAAGGAT 3586  
 |||||  
 Db 297 CTTAATGATGATGCTGTAACATGCAATATACGAGAAATATTGTAATGTCAAGGAT 238  
 |||||  
 QY 3587 TGTGTTCTGTGATTTTCCAAAGTATCCGATGCGCAAGAGGTTGAACACCATGTAGAG 3646  
 |||||  
 Db 237 TGTGTTTGGATATGTCGAAGTCGGTCTCTTCGAGAGAATCTGAGACGACATTGAAA 178  
 |||||  
 QY 3647 CCAGTTTTCGTACCGCGCGGACCGCAAGGCTGCAGGACTACTCGAANAATCTGTT 3706  
 |||||  
 Db 177 CCGTGATCAGGACTGCTGCTGAAACCTCGAACAACCTGGAATTTGCGNAACCTTGTC 118  
 |||||  
 QY 3707 GCAATGATTAAAGAAATTTCAAGCAGACGACCTGACGGGACGATTGACATTGAGAGC 3766  
 |||||  
 Db 117 GCGATGATCAAGAAATTTCACTCTCCGAAATTAATAGGGTCTGTTGACATCGAAGAC 58  
 |||||  
 QY 3767 ACCGATCTGTGTAGTAGAATGTTTTCATAGCTATTTTATTAATAA 3816  
 |||||  
 Db 57 ACCGCTCTCTAGTAGTAGAATGTTTTCATGATCATCTTTTATTAATAA 8  
 |||||

RESULT 5  
 BM068523/c 668 bp mRNA linear EST 11-SEP-2002  
 LOCUS KS08025D04 KS08 Capsicum annuum cDNA, mRNA sequence.

DEFINITION BM068523  
 ACCESSION BM068523  
 VERSION BM068523.1 GI:22788643

KEYWORDS EST.  
 SOURCE Capsicum annuum  
 ORGANISM Capsicum annuum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 asterids; lamids; Solanales; Solanaceae; Capsicum.

REFERENCE 1 (bases 1 to 668)  
 AUTHORS Lee, S., Kim, S.-Y., Chung, Y.-H., Shin, H.-J., Goh, S.-H., Pai, H.-S., Hur, C.-G. and Choi, D.  
 TITLE Generation of Expressed Sequence Tags from Hot Pepper (Capsicum annuum L.) and Sequence Analysis in Relation to Hypersensitive Response Against Pathogen

JOURNAL Unpublished (2001)  
 COMMENT Contact: Doil Choi  
 Genome Research Center and National Center for Genome Information  
 Korea Research Institute of Bioscience and Biotechnology  
 P.O. Box 115, Yuseong, Taejeon, 305-600, Republic of Korea  
 Tel: 82-42-860-4340

FEATURES  
 source  
 1. 668  
 /organism="Capsicum annuum"  
 /mol\_type="mRNA"  
 /cultivar="Hang Keun"  
 /db\_xref="taxon:4072"  
 /tissue type="anther"  
 /dev stages="10 weeks after germination"  
 /clone lib="KS08"  
 /note="Vector: pBluescript SK(-)"

ORIGIN  
 Query Match 4.7%; Score 297.6; DB 4; Length 668;  
 Best Local Similarity 65.9%; Pred. No. 7.7e-62;  
 Matches 432; Conservative 0; Mismatches 224; Indels 0; Gaps 0;

QY 3159 ATGTGAACACCGTTTCATGAGATCCAAAGAGAAACCTTTGAAGATGTGCTGTCAGAT 3218

Db 663 ATGTGCAATACCGTTTCATGAGTGCAGGGGAAACGTTTGAAGACGTTCTCATTAGTGAGAC 604  
 |||||  
 QY 3219 TGACGGCAATCCCACTGACTCTGATTTTCCAAGTCTTCCCGCATGTTCTAGTCGCTCTGA 3278  
 |||||  
 Db 603 TGACGCCCTACACCCGTTGGGAATTAATTTCAAGCAGAGTCCGACCTGTTGCTCTGTTGT 544  
 |||||  
 QY 3279 CTAGACACACAAAGAGCTTCAAAATATTACACCGTAGTGTGTAGATCCTTTAGTACAGATAA 3338  
 |||||  
 Db 543 CTAGGCATACAAGGTCAATCAAAATATTACACTGTTGTACTAGATGCACTGTTTCAGTGC 484  
 |||||  
 QY 3339 TTAGTGATTTGCTCTTTTAAGCTCTCTTTTAGAAATGATATGTTAGTAGAGGTA 3398  
 |||||  
 Db 483 TTAGAGATTTGGAGTGTGTGAGTAGTACCTGTTTAGATATGTACAAAGTGTGATGTGCGA 424  
 |||||  
 QY 3399 GTAGATAGCAATTAACAGATGAGTGTTCAAAGGTCAATACTCTCTTTGTGCAACAC 3458  
 |||||  
 Db 423 CTCAATAGCAATTAACAGATAGATCGGTGTACAAGGTGTTAACCTTTTCGTCGAGGCC 364  
 |||||  
 QY 3459 CTAAATCAGGAGACTTTCCAGATCTACAGTCTTATTACGATGTATGCTCCCTCGTATAA 3518  
 |||||  
 Db 363 CGAAACAGGAGATGTTTCTGACATGCAATATTACTATGACAAGTGTTCGCGGAAACA 304  
 |||||  
 QY 3519 GTACTATCTTAAACAAGTATGATCTGTTACCATGAGGTTACGTTGATAATAGTCTTATG 3578  
 |||||  
 Db 303 GTACTATCTTAAATGATGATGCTGTAATGCAATATGCAAAATACGAGAAATTAATTTGAATG 244  
 |||||  
 QY 3579 TGAAGGATGTGTTCTTGTGATTTTCCAAAGATTTCCGATGTCGCAAGAGGTGAACCAT 3638  
 |||||  
 Db 243 TCAAGGATGTGTTGATGATGTCGAGTGTGCTCTTCCGAGAGATCTCGAGGA 184  
 |||||  
 QY 3639 GTCTAGAGCCAGTTTTCGTACCGCGCGGACCGCAAGGGCTGCAGGACTACTCGAAA 3698  
 |||||  
 Db 183 CATTGAAACCTGTGATCAGGACTGCTGCTGAAACCTCGAAGACCTGGATTGTGGAAA 124  
 |||||  
 QY 3699 ATCTGTTGCAATGATTAAGAAATTTCAAGCAGCAGACCTGACGGGAGGATTGACA 3758  
 |||||  
 Db 123 ACTTGGTGGGATGATCAAAAGAAATTTCAACTCTCCGAAATTAATAGGGGTCTGTGACA 64  
 |||||  
 QY 3759 TTGAGACACCGCATCTGTTGTAGTAGATAAGTTTTTTTGTAGTCTATTTTATTA 3814  
 |||||  
 Db 63 TCGAAGACACCGCTCTCTAGTAGATAGATAGTTTTTTTGTAGTCTATTTATTA 8  
 |||||

RESULT 6  
 BM068294/c 663 bp mRNA linear EST 11-SEP-2002  
 LOCUS KS08020D04 KS08 Capsicum annuum cDNA, mRNA sequence.

DEFINITION BM068294  
 ACCESSION BM068294  
 VERSION BM068294.1 GI:22788414

KEYWORDS EST.  
 SOURCE Capsicum annuum  
 ORGANISM Capsicum annuum

REFERENCE 1 (bases 1 to 663)  
 AUTHORS Lee, S., Kim, S.-Y., Chung, Y.-H., Shin, H.-J., Goh, S.-H., Pai, H.-S., Hur, C.-G. and Choi, D.  
 TITLE Generation of Expressed Sequence Tags from Hot Pepper (Capsicum annuum L.) and Sequence Analysis in Relation to Hypersensitive Response Against Pathogen

JOURNAL Unpublished (2001)  
 COMMENT Contact: Doil Choi

Genome Research Center and National Center for Genome Information  
 Korea Research Institute of Bioscience and Biotechnology  
 P.O. Box 115, Yuseong, Taejeon, 305-600, Republic of Korea  
 Tel: 82-42-860-4340  
 Fax: 82-42-860-4309  
 Email: doil@mail.kribb.re.kr  
 High quality sequence stop: 663.  
 Location/Qualifiers  
 1. 663

FEATURES  
 source  
 1. 663

```

/organism="Capsicum annuum"
/mol_type="mRNA"
/cultivar="Hang Keun"
/db_xref="taxon:4072"
/tissue_type="anther"
/dev_stages="10 weeks after germination"
/clone_lib="KS08"
/note="Vector: pBluescript SK(-)"

ORIGIN
Query Match      4.7%; Score 296.6; DB 4; Length 663;
Best Local Similarity 65.8%; Pred. No. 1.4e-61;
Matches 431; Conservative 0; Mismatches 224; Indels 0; Gaps 0;

QY 3163 GAACACCGTTCATGAGATCCAAAGGAGAAACCTTTGAAGATGTCGCTGGTCAGATTGAC 3222
DB 663 GCATACCGTTCATGAGTGCAGAGGAGAAACGTTTGAAGACGCTCTCATTAGTGAGACTGAC 604

QY 3223 GGCACACTCCAGTCTGATTTCCAAAGTCTCCCGCATGTTCTAGTCGCTCTGACTAG 3282
DB 603 GCGTACACCGTGGGAAATTTCAAGACAGAGTCGACACCTGTTGGTCTCGTCTCTAG 544

QY 3283 ACACACAAAGAGCTTCAATATATTACACCGTAGTGTAGATCTTTAGTACAGATAATTAG 3342
DB 543 GCATACAGGTCAATCAATATATTACATGTTGTACTAGTGCAGTCTTTCAAGTCTTAG 484

QY 3343 TGAATTTGCTCTTTAAGCTCTCTCTTTTGAAGATGATATGAGGAGAGGAGTAGTAG 3402
DB 483 AGATTTGGAGTGTGAGTAGTACCTGTTAGATATGTACAAAGTTGATGTGTCGACTCA 424

QY 3403 ATAGCAATTTACAGATGATGATGCTTCAAGGTCAATATCTTTTGGGCAACCTTAA 3462
DB 423 ATAGCAATTTACAGATGATGATGCTTCAAGGTGTAACAAAGGTGTTAACTTTTGGTGGCAGCCCGAA 364

QY 3463 ATCAGGAGACTTTCAGATCTACAGTCTTATTAACGATGATGCTTCCCTGTTGTAATAGTAC 3522
DB 363 AACAGAGATGTTTCTGACATCAATATTAATCTATGACAGTGTTCGCGGAGAAACAGTAC 304

QY 3523 TATACCTTAAAGATGATGCTGTTTCAAGAGGTTACGTTGATATAGTCTTTAATGTGAA 3582
DB 303 TATACCTTAAAGATGATGCTGTTTCAAGAGGTTACGTTGATATAGTCTTTAATGTGAA 244

QY 3583 GGATTTGCTTCTGATTTTCCAAAGTATTCGATGTCGAAAGAGGTTGAACCATGCT 3642
DB 243 GGATTTGCTTCTGATTTTCCAAAGTATTCGATGTCGAAAGAGGTTGAACCATGCT 184

QY 3643 AGAGCCAGTTTTCGCTACCGCGCGGAAACCGCAAGGGCTGCGAGGACTACTCGAAATCT 3702
DB 183 GAAACCTGTGATCAGACTGCTGCTGAAACCTCGAAACCTGGAATGTTGGAAACTT 124

QY 3703 GGTTCGAATGATTAAGAAATTTCAAGCAGCAGACCTGAGCGGAGGAGTGAATGAA 3762
DB 123 GGTTCGAATGATTAAGAAATTTCAAGCAGCAGACCTGAGCGGAGGAGTGAATGAA 64

QY 3763 GAGCAGCCAGTCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3817
DB 63 AGACACCGCTCTCTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAG 9

RESULT 7
BM068316/c BM068316 663 bp mRNA linear EST 11-SEP-2002
LOCUS KS08020F12 KS08 Capsicum annuum cDNA, mRNA sequence.
DEFINITION BM068316
ACCESSION BM068316
VERSION BM068316.1 GI:22788436
KEYWORDS EST.
SOURCE Capsicum annuum
ORGANISM Capsicum annuum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamiales; Solanales; Solanaceae; Capsicum.
REFERENCE 1 (bases 1 to 663)
AUTHORS Lee,S., Kim,S.-Y., Chung,Y.-H., Shin,H.-J., Goh,S.-H., Pai,H.-S.,

```

```

Hur,C.-G. and Choi,D.
Generation of Expressed Sequence Tags from Hot Pepper (Capsicum
annuum L.) and Sequence Analysis in Relation to Hypersensitive
Response Against Pathogen
Unpublished (2001)
Contact: Doil Choi
Genome Research Center and National Center for Genome Information
Korea Research Institute of Bioscience and Biotechnology
P.O. Box 115, Yuseong, Taejeon, 305-600, Republic of Korea
Tel: 82-42-860-4340
Fax: 82-42-860-4309
Email: doil@mail.kribb.re.kr
High quality sequence stop: 663.
Location/Qualifiers
1..663
/organism="Capsicum annuum"
/mol_type="mRNA"
/cultivar="Hang Keun"
/db_xref="taxon:4072"
/tissue_type="anther"
/dev_stages="10 weeks after germination"
/clone_lib="KS08"
/note="Vector: pBluescript SK(-)"

ORIGIN
Query Match      4.7%; Score 296.2; DB 4; Length 663;
Best Local Similarity 65.8%; Pred. No. 1.7e-61;
Matches 430; Conservative 0; Mismatches 223; Indels 0; Gaps 0;

QY 3165 ACACCGTTCATGAGATCCAAAGGAGAAACCTTTGAAGATGTCGCTGGTCAGATTGACGG 3224
DB 661 ATACCGTTCATGAGTGCAGGAGGAAACGTTTGAAGACGCTCTCATTAGTGAGACTGACGC 602

QY 3225 CAATCCACTGACTCTGATTTTCCAAAGTCTTCCCGCATGTTCTAGTCTCTGACTAGAC 3284
DB 601 CTACACCGTGGGAAATTTTCAAAGCAGAGTCCGACCTGTTGCTCTGCTGTCTAGGC 542

QY 3285 ACACAAGAGCTTCAATATATTACACCGTAGTGTAGATCTTTAGTACAGATAATTAGTG 3344
DB 541 ATACAAGGTCAATCAATATATTACATGTTGACTAGATGCAGTCTGTTTCAAGTCTTAGAG 482

QY 3345 ATTTGCTCTTTTAAAGTCTTCTTTTAAAGATGTTATATGTTAGAAATGTTATGTTAGAACGAGTGTAGAT 3404
DB 481 ATTTGGAGTGTGAGTAGTTACCTGTTAGATATGTTACAAAGTGTGATGTGCTGCTCAAT 422

QY 3405 AGCAATTAACAGATGATGATGTTTCAAGGTCAATATCTTTTGTGCAACACCTAAAT 3464
DB 421 AGCAATTAACAGATGATGATGTTTCAAGGTGTTAACTTTTCTGCGAGCCCGGAAAA 362

QY 3465 CAGGAGACTTTCAGATCTACAGTCTTATACGATGATGCTCTCTCTGTTAGTACTA 3524
DB 361 CAGGAGATGTTTCTGACATGCAATATTAATGATGCAAGTGTTCGCGGAAACAGTACTA 302

QY 3525 TACTTAAACAGTATGATGCTGTTTACCATGAGTTACGTTAGTAATAAGTCTTAAATGTGAAGG 3584
DB 301 TACTTAAAGTATGATGCTGTTTACCATGAGTTACGTTAGTAATAAGTCTTAAATGTGAAGG 242

QY 3585 ATTGTTTCTTGAATTTTCCAAAGATTTCCGATGTCGCAAGAGGTTGAACCATGCTAG 3644
DB 241 ATTGTTTCTTGAATTTTCCAAAGATTTCCGATGTCGCAAGAGGTTGAACCATGCTAG 182

QY 3645 AGCCAGTTTTCGTTACCGCGCGGAAACCGCAAGGGCTGACGAGTACTCTGAAATCTGG 3704
DB 181 AACCTGTGATCAGGACTGCTGCTGAAACCTCGAAAAACCTGGATTTGTTGAAAAAATTGG 122

QY 3705 TTGCAATGATTAAGAAATTTTCAACGCAACAGACCTGACCGGGACGATTGACATTGAGA 3764
DB 121 TCGCATGATCAAAAGAAATTTTCAACTCTCCGAAATTAATAGGGGTCTGTGACATCGAAG 62

QY 3765 GCACCGCATCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3817
DB 61 ACACCGCTCTCTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAG 9

```

```

RESULT 8
BM067993/c
LOCUS      BM067993      701 bp      mRNA      linear      EST 11-SEP-2002
DEFINITION KS08015C03 KS08 Capsicum annuum cDNA, mRNA sequence.
ACCESSION  BM067993
VERSION     BM067993.1  GI:22788113
KEYWORDS    EST.
SOURCE      Capsicum annuum
ORGANISM    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; lamids; Solanales; Solanaceae; Capsicum.
REFERENCE   1 (bases 1 to 701)
AUTHORS     Lee,S., Kim,S.-Y., Chung,Y.-H., Shin,H.-J., Goh,S.-H., Pai,H.-S.,
            Hur,C.-G. and Choi,D.
TITLE       Generation of Expressed Sequence Tags from Hot Pepper (Capsicum
            annuum L.) and Sequence Analysis in Relation to Hypersensitive
            Response Against Pathogen
JOURNAL     Unpublished (2001)
COMMENT     Contact: Doil Choi
            Genome Research Center and National Center for Genome Information
            Korea Research Institute of Bioscience and Biotechnology
            P.O. Box 115, Yuseong, Taejeon, 305-600, Republic of Korea
            Tel: 82-42-860-4340
            Fax: 82-42-860-4309
            Email: doil@mail.kribb.re.kr
            High quality sequence stop: 701.

FEATURES             source
     source
     1..701
     /organism="Capsicum annuum"
     /mol_type="mRNA"
     /cultivar="Hang Keun"
     /db_xref="taxon:4072"
     /tissue_type="anther"
     /dev_stage="10 weeks after germination"
     /clone_lib="KS08"
     /note="Vector: pBluescript SK(-)"

ORIGIN
Query Match      4.7%; Score 295.8; DB 4; Length 701;
Best Local Similarity 65.2%; Pred. No. 2.2e-61;
Matches 435; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 3146 AAGGCTATAAAGTGTGAACACCGTTCATGAGATCCAAAGGAGAAACCTTTGAAGATGTG 3205
DB 674 AGGGGTTACGAAGAGTACATACCGTTCATGAGGTGCAAGGGGAAACGTTGAAGACGTC 615
QY 3206 TCGCTGGTCAGATTGACGGCACTCCACTGACTCTGATTTCCAAAGTCTTCCCGCATGTT 3265
DB 614 TCATTAGTGAGACTGACGCCCTACACCCGTTGGGAATAATTTCAAAGCAGAGTCCGCACCTG 555
QY 3266 CTAGTCGCTCTGACTAGACACAAAGAGTTCAAATATTACCGTAGTGTAGATCCT 3325
DB 554 TTGGTCTCGTGTCTAGGCATACAAAGTCAATCAATTAATTAACCTGTTGACTAGATGCA 495
QY 3326 TTAGTACAGATAAATPAGTGTGTTCTTTAAAGCTCCTTCCCTTTAGAAATGTATATG 3385
DB 494 GTCGTTTCAGTGTCTAGAGATTGGAGTGTGTGAGTAGTTACCTGTTAGATATGTACAAA 435
QY 3386 GTAGNAGCAGGTAGTAGATAGCAATACAGATGATGAGTTCAGAGTTCAAAGTCAATATCTC 3445
DB 434 GTTGATGTGCTGACTCAATAGCAATACAGATGATGAGTTCAGAGTTCAGAGTTCAGAGTTC 375
QY 3446 TTGTGGCAACACCTTAATCAGGAGCTTCCAGATCTACAGTCTTATTACGATGTATGC 3505
DB 374 TTGCTCGCAGCCCGAAACAGGAGATGTTTCTGACATGCAATATTACTATGACAAAGTGT 315
QY 3506 CTCCTGTGTATAGTACTATATCTTAACAAGTATGATGCTGTGTTACATGAGTTACGTGAT 3565
DB 314 TTGCGGGAAACAGTACTATATCTTAATGAGTATGATGCTGTAACTATGCAATACAGAA 255
QY 3566 AATAGTCTTAATGTGAAGGATTGTCTTCTTGAATTTTCCAAAGATTTCCGATGCCAAG 3625

```

---

```

Db 254 AATAATTGAATGTCAAGGATTGTGTGTTGATATGTCGAAGTCGGTGCTCTTCCGAGA 195
QY 3626 GAGGTGAACACATCTCTAGAGCCAGTTTTCGGTACCGGGGGAACCGCCAGGGCTGCA 3685
DB 194 GAATCTGAGACGACATTGAAACCTGTGATCAGGACTCTGCTGAAAAACCTTCGAAAACT 135
QY 3686 GGACTACTCGAAAAATCTGGTTGCAATGATTAAAAAGAAATTTCAACGACACAGACTGACG 3745
DB 134 GGATGTTGGAAAACTTGGTCGCGATGATCAAAAGAAATTTCAACTCTCCGAAATTAATA 75
QY 3746 GGGAGCAATTGACATGAGAGACCGCATCTGTTCTAGTAGATAAGTTTTCATAGTAGCTAT 3805
DB 74 GGGTCGTTGACATCGAAGACACCGCTTCTCTAGTAGTAGATAAGTTTTCATAGTAGCTAT 15
QY 3806 TTTATTA 3812
DB 14 TTTATTA 8

RESULT 9
BM067343/c
LOCUS      BM067343      665 bp      mRNA      linear      EST 11-SEP-2002
DEFINITION KS08003C07 KS08 Capsicum annuum cDNA, mRNA sequence.
ACCESSION  BM067343
VERSION     BM067343.1  GI:22787463
KEYWORDS    EST.
SOURCE      Capsicum annuum
ORGANISM    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; lamids; Solanales; Solanaceae; Capsicum.
REFERENCE   1 (bases 1 to 665)
AUTHORS     Lee,S., Kim,S.-Y., Chung,Y.-H., Shin,H.-J., Goh,S.-H., Pai,H.-S.,
            Hur,C.-G. and Choi,D.
TITLE       Generation of Expressed Sequence Tags from Hot Pepper (Capsicum
            annuum L.) and Sequence Analysis in Relation to Hypersensitive
            Response Against Pathogen
JOURNAL     Unpublished (2001)
COMMENT     Contact: Doil Choi
            Genome Research Center and National Center for Genome Information
            Korea Research Institute of Bioscience and Biotechnology
            P.O. Box 115, Yuseong, Taejeon, 305-600, Republic of Korea
            Tel: 82-42-860-4340
            Fax: 82-42-860-4309
            Email: doil@mail.kribb.re.kr
            High quality sequence stop: 665.

FEATURES             source
     source
     1..665
     /organism="Capsicum annuum"
     /mol_type="mRNA"
     /cultivar="Hang Keun"
     /db_xref="taxon:4072"
     /tissue_type="anther"
     /dev_stage="10 weeks after germination"
     /clone_lib="KS08"
     /note="Vector: pBluescript SK(-)"

ORIGIN
Query Match      4.7%; Score 295.6; DB 4; Length 665;
Best Local Similarity 65.7%; Pred. No. 2.4e-61;
Matches 430; Conservative 0; Mismatches 224; Indels 0; Gaps 0;

QY 3161 GTGAACACCGTTCATGAGATCCAAAGGAGAAACCTTTGAAGATGTGTCGCTGGTCAGATTG 3220
DB 665 GTGCATACCGTTCATGAGGTGCAAGGGGAAACGTTTGAAGACGCTCTCATAGTGAGACTG 606
QY 3221 ACGGCAATCCACTGACTCTGATTTCCAAAGTCTTCCCGCATGTTCTAGTCGCTGCTGACT 3280
DB 605 ACGCCTACACCGCTGGGAATAATTTCAAAGCAGAGTCCGCACCTGTGTGGTCTCGTTGCT 546
QY 3281 AGACACAAAGAGCTTCAATATATTACCGTAGTGTAGATCCTTTAGTACAGATAATT 3340

```





Tel: 82-42-860-4340  
Fax: 82-42-860-4309  
Email: doilemail.kribb.re.kr  
High quality sequence stop: 693.

## FEATURES

source  
1. .693  
/organism="Capsicum annuum"  
/mol\_type="mRNA"  
/cultivar="Hang Keun"  
/db\_xref="taxon:4072"  
/tissue\_type="anther"  
/dev\_stage="10 weeks after germination"  
/clone\_lib="KS08"  
/note="Vector: pBluescript SK(-)"

## ORIGIN

Query Match 4.6%; Score 293.6; DB 4; Length 693;  
Best Local Similarity 64.3%; Pred. No. 7.5e-61;  
Matches 440; Conservative 0; Mismatches 244; Indels 0; Gaps 0;  
QY 3113 TTCACTCAGGCTGATAAATTTGAGTTAGAGGAGAGGCTATAAGAAATGTGAACACCGTT 3172  
Db |||||  
QY 692 TTCACTCAATCAGACAGATCTGCTCTCAAGGGTTACGAAGATGTGCTACTGTT 633  
Db |||||  
QY 3173 CATGAGATCCAGGAGAAACCTTTGAAGATGTCGCTGGTCAGATTCACGGCAACTCCA 3232  
Db |||||  
QY 632 CATGAGGTGCAAGGGGAAACGTTTGAAGACGCTCTCATTAGTGAGACTGACGCTACACCC 573  
QY 3233 CTGACTCTGATTTCCAAAGCTTCCCGGATGTTCTAGTCGCTCTGACTAGACACAAAG 3292  
Db |||||  
QY 572 GTGGGAATAATTTCAAGACAGAGTCGCGACCTGTTGGTCTCGTTGCTAGGATACAAAG 513  
QY 3293 AGCTTTCAATATTTACACCGTAGTGTGTAGATCCCTTTAGTACAGATAAATAGTGTGCT 3352  
Db |||||  
QY 512 TCAATCAATATTTACACTGTTGTACTAGATGACGTCGTTTCAGTCTTAGAGATTTGAG 453  
QY 3353 TCTTTAAGCTCTCTCTTTGAATATGATATGTTAGAGAGAGGTTAGATAGCAATTA 3412  
Db |||||  
QY 452 TGTGTGAGTAGTTACCTGTAGATATGACAAAGTTGATGTCTGACTCAATAGCAATTA 393  
QY 3413 CAGATGGATGCGATGTTCAAGAGTCATAATCTCTTTTGGCAACACCTTAATCAGGAGAC 3472  
Db |||||  
QY 392 CAGATAGATCGGTGTACAAAGGTGTTAACTTTTCGCGAGCCCGGAAACAGGAGAT 333  
QY 3473 TTTCAGATCTACAGTCTCTATTACAGATGATGCTCTCCCTGGTAAATAGTACTTAAAC 3532  
Db |||||  
QY 332 GTTCTGACATGCAATATTTACTATGACAAAGTGTGTCGCGGAAACAGTACTATCTTAAT 273  
QY 3533 AAGTATGATGCTGTACCATGAGTTAGTGATATAGTCTTTAATGTGAAGATTTGTT 3592  
Db |||||  
QY 272 GAGTATGATGCTGTAACTATGCAATACGAGAAATATAATTTGAATGTCAAGGATTTGTG 213  
QY 3593 CTGTATTTTCCAAAGTATTCGAGATGCCAAGAGGTTGAAACCATGCTAGAGCCAGTT 3652  
Db |||||  
QY 212 TTGATATGTCGAATCGGTGCTCTTCGAGAGATCTGAGACACATTTGAACCTGTG 153  
QY 3653 TTGCGTACCGGGGGAACCGCAAGGCTCGAGACTACTCGAATAATCTGGTTGCAATG 3712  
Db |||||  
QY 152 ATCAGGACTGCTGTGAAACAACTCGAATACTGGAATTTGGAATACTTGGTCGCGATG 93  
QY 3713 ATTTAAAGAAATTTCAACGACCGACCTGACGGGACGATTGACATTTGAGACACCGCA 3772  
Db |||||  
QY 92 ATCAAAAGAAATTTCACTCTCCCGAATTAATAGGGGTCGTTGATCGAAGACACCGCT 33  
QY 3773 TCTGTGTAGTAGATAAGTTTTTT 3796  
Db |||||  
QY 32 TCTCTAGTAGTAGATAAGTTTTTT 9

RESULT 12  
BM068536/c  
LOCUS  
DEFINITION KS08025F01 KS08 Capsicum annuum cDNA, mRNA EST 11-SEP-2002  
556 bp linear  
656 bp mRNA  
556 bp linear  
EST 11-SEP-2002

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

BM068536  
BM068536.1 GI:22788656  
EST.  
Capsicum annuum  
Capsicum annuum

REFERENCE  
AUTHORS

Eukaryota, Viridiplantae, Streptophyta, Embryophyta, Tracheophyta,  
Spermatophyta, Magnoliophyta, eudicotyledons, core eudicots,  
asterids, lamids, Solanales, Solanaceae, Capsicum.  
1 (bases 1 to 656)  
Lee, S., Kim, S.-Y., Chung, Y.-H., Shin, H.-J., Goh, S.-H., Pai, H.-S.,  
Hur, C.-G. and Choi, D.

## TITLE

Generation of Expressed Sequence Tags from Hot Pepper (Capsicum  
annuum L.) and Sequence Analysis in Relation to Hypersensitive  
Response Against Pathogen

JOURNAL  
COMMENT

Unpublished (2001)  
Contact: Doll Choi  
Genome Research Center and National Center for Genome Information  
Korea Research Institute of Bioscience and Biotechnology  
P.O. Box 115, Yusong, Taejeon, 305-600, Republic of Korea  
Tel: 82-42-860-4340  
Fax: 82-42-860-4309  
Email: doilemail.kribb.re.kr  
High quality sequence stop: 656.

FEATURES  
source

1. .656  
Location/Qualifiers  
/organism="Capsicum annuum"  
/mol\_type="mRNA"  
/cultivar="Hang Keun"  
/db\_xref="taxon:4072"  
/tissue\_type="anther"  
/dev\_stage="10 weeks after germination"  
/clone\_lib="KS08"  
/note="Vector: pBluescript SK(-)"

## ORIGIN

Query Match 4.6%; Score 292.8; DB 4; Length 656;  
Best Local Similarity 65.7%; Pred. No. 1.2e-60;  
Matches 426; Conservative 0; Mismatches 222; Indels 0; Gaps 0;  
QY 3170 GTTCATGAGATCCAAAGGAGAACTTTGAAGATGTCGCTGCTCAGATTGACGCAACT 3229  
Db |||||  
QY 656 GTTCATGAGTGCAGGGGAAACGTTTGAAGACGTTCTATTAGTAGACTGACGCTTACA 597  
Db |||||  
QY 3230 CCACCTGACTCTGATTTTCCAAAGTCTTCCCGCATGTTCTAGTCGCTCTGACTAGACACACA 3289  
Db |||||  
QY 596 CCGTGGGAAATATTTCAAAGCAGAGTCGCGACCTGTTGCTGCTGCTGCTAGCATACA 537  
QY 3290 AAGAGCTTCAATATTTACACCGTAGTGTAGATCTTTTAGTACAGATAAATTAGTATTG 3349  
Db |||||  
QY 536 AGGTCAATCAAAATTTTACACTGTTTACTAGATGCACTGCTTTTCAAGTCTTAGAGATTG 477  
QY 3350 TCTTCTTTAAGCTCTTCTCTTTTAGAAATGTATATGTAGAAGCAGGTAGTAGTAGCAA 3409  
Db |||||  
QY 476 GAGTGTGTAGTAGTACCTGTTAGATATGTACAAAGTTGATGTGCGACTCAATAGCAA 417  
QY 3410 TTACAGATGATCAGGTGTTCAAAGGTCAATATCTCTTTTGGCAACACCTTAATCAGGA 3469  
Db |||||  
QY 416 TTACAGATGAAATCGGTGTACAAAGGTGTTAACTTTTCTGTCGAGCCCGGAAACAGGA 357  
QY 3470 GACTTTCCAGATCTACAGTCTTATTAGATGATGCTCCCTGGTAAATAGTACTATCTT 3529  
Db |||||  
QY 356 GATGTTTCTGACATGCAATATTTACTATGACAGTGTTCGCGGAAACAGTACTATCTT 297  
QY 3530 AACAAAGTATGATGCTGTACCATGAGTTACGTGATAAATAGTCTTAAATGTGAAGGATTGT 3589  
Db |||||  
QY 296 AATGATATGATGCTGTACTATGCAATACGAGAAATAATTTGAATGTCAAGGATTGT 237  
QY 3590 GTTCTTGAATTTTCCAAAAGTATTTCCGATGCGCAAGAGGTGAAACCATGCTTAGAGCCA 3649  
Db |||||  
QY 236 GTGTTGGATATGTGGAAGTCGGTGCCTCTTCCGAGAGAAATCTGAGACGACATTTGAAACCT 177  
QY 3650 GTTTTGGTACCGCGCGGAAACCGCAAGGCTGCGAGGACTACTCTGAAATCTGTTTGA 3709  
Db |||||



```
QY 3206 TCGTGTGTCAGATTGACGGCACTCCACTGACTCTGATTTCAGATCTTCCCGCATGTT 3265
Db |||||
QY 608 TCATTAGTGAGACTGACCGCTACACCCGTGGGAATAATTTCAAAGCAGAGTCCGCACCTG 549
Db |||||
QY 3266 CTAGTCGCTCTGACTAGACACACAAAGAGCTTCAAAATATTACCGTAGTGTAGATCCT 3325
Db |||||
QY 548 TTGGTCTCGTTGCTTAGGCATACAAAGGTCATCAAAATATTACACTGTTGTACTAGATGCA 489
Db |||||
QY 3326 TTAGTACAGATAAATTAGTGATTGTTCTTTTAAGCTCTTCCCTTTTGAAGATGATATG 3385
Db |||||
QY 488 GTCGTTTCAGTCTTAGAGATTGGAGTGTGAGTAGTTACCTGTTAGATATGTACAAA 429
Db |||||
QY 3386 GTAGAAGCAGGTAGTAGCAATTAAGATGAGTGTTCAAAGGTCATAATCTC 3445
Db |||||
QY 428 GTTGATGTGTCGACTCAATAGCAATTAAGATGAGTGTTCGTTACAAAGGTTTAAACCTT 369
Db |||||
QY 3446 TTTGTGGCAACACCTTAATCAGGAGCTTCCAGATCTACAGTCTTATTAAGATGTATGC 3505
Db |||||
QY 368 TTGTCGCGAGCCCGCAACAGGAGATGTTTCTGACATGCAATATTACTATGCAAGTGT 309
Db |||||
QY 3506 CTCCCTGTGTAATAGTACTATCTTAAACAAGTATGATGCTGTACCATGAGGTACGTGAT 3565
Db |||||
QY 308 TTGCGGGAACAGTACTATCTTAAATGAGTATGATGCTGTAACTATGCAATACGAGAA 249
Db |||||
QY 3566 AATAGCTTAATGGAAGATGTTGTTCTGATTTTCCAAAGATTTCCGATGCCAAAG 3625
Db |||||
QY 248 AATAATTTGAATGTCAAGGATTTGTTGATATGTCGAAGTGGTCTCTTCCGAGA 189
Db |||||
QY 3626 GAGGTGAACCATGCTAGAGCCAGTTTTCGTCACCGCGGACCCGCAAGGGCTGCA 3685
Db |||||
QY 188 GAATCTGAGACGACATTTGAACCTGTGATAGGACTGCTGTGAAACACCTCGAAACCT 129
Db |||||
QY 3686 GGACTACTCGAAATCTGTTGCAATGATTAAGAAATTTCAACGACACGACCTGACG 3745
Db |||||
QY 128 GGATGTTGGAATCTGTCGCGATGATCAAGAAATTTCAACTCTCCGAAATTAATA 69
Db |||||
QY 3746 GGGAGATGATGACATGAGAGCAGCCGATCTGTTGTAGTAGATAGTGTATTTTGTAGCTA 3804
Db |||||
QY 68 GGGGTGCTGTGACATCGAAGACACCGCTTCTCTAGTAGTAGATAGTATTTTGTATGCATA 10
Db |||||
```

## RESULT 15

BM067465/c  
LOCUS KS08005D10 KS08 Capsicum annum cDNA, mRNA linear EST 11-SEP-2002  
DEFINITION BM067465  
ACCESSION BM067465  
VERSION BM067465.1 GI:22787585  
KEYWORDS EST.

## SOURCE

Capsicum annum  
Capsicum annum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
asterids; lamiales; Solanales; Solanaceae; Capsicum.  
1 (bases 1 to 679)

## REFERENCE

AUTHORS Lee, S.-Y., Kim, S.-Y., Chung, Y.-H., Shin, H.-J., Goh, S.-H., Pai, H.-S.,  
Hur, C.-G. and Choi, D.

## TITLE

Generation of Expressed Sequence Tags from Hot Pepper (Capsicum  
annuum L.) and Sequence Analysis in Relation to Hypersensitive

## JOURNAL

COMMENT Response Against Pathogen  
Unpublished (2001)  
Contact: Doil Choi  
Genome Research Center and National Center for Genome Information  
Korea Research Institute of Bioscience and Biotechnology  
P.O. Box 115, Yuseong, Taejeon, 305-600, Republic of Korea  
Tel: 82-42-860-4340  
Fax: 82-42-860-4309  
Email: doil@mail.kribb.re.kr

High quality sequence stop: 679.

## FEATURES

Location/Qualifiers  
1..679  
/organism="Capsicum annum"  
/mol\_type="mRNA"  
/cultivar="Hang Keun"

/db\_xref="taxon:4072"  
/tissue\_type="anther"  
/dev\_stage="10 weeks after germination"  
/clone\_lib="KS08"  
/note="Vector: pBluescript SK(-)"

## ORIGIN

Query Match 4.6%; Score 292.4; DB 4; Length 679;  
Best Local Similarity 65.4%; Pred. No. 1.5e-60;  
Matches 428; Conservative 0; Mismatches 226; Indels 0; Gaps 0;  
QY 3146 AAGGGCTATAGAGATGCAACACCGTTTCATGAGATCCAAAGGAGAACTTTGAAGATGTG 3205  
Db |||||  
QY 664 AAGGGTTTACGAAGATGTGCATACCGTTTCATGAGTGCAAGGGGAAACGTTTGAAGACGTC 605  
Db |||||  
QY 3206 TCGTGTGTCAGATTGACGGCAACTCCACTGACTCTGATTTTCCAAAGTCTTCCCCCATGTT 3265  
Db |||||  
QY 604 TCATTAGTGAGACTGACGCTACACCGTGGGAATAATTTCAAAGCAGAGTCCGCACCTG 545  
Db |||||  
QY 3266 CTAGTCGCTCTGACTAGACACACAAAGAGCTTCAAAATATTACACCGTAGTGTAGATCCT 3325  
Db |||||  
QY 544 TTGCTCTCGTTGCTTAGGCATACAAAGTCAATCAAAATATTACACTGTTGTACTAGATGCA 485  
Db |||||  
QY 3326 TTAGTACAGATAAATTAGTGATTGTTCTTTTAAGCTCTTCCCTTTTGAAGATGTATATG 3385  
Db |||||  
QY 484 GTCGTTTTCAGTCTTAGAGATTTGGAGTGTGTAGTAGTTACCTGTTAGATATGTACAAA 425  
Db |||||  
QY 3386 GTAGAAGCAGGTAGTACATGAGCAATTAAGATGATGATGATGTTTCAAAGGTCATAATCTC 3445  
Db |||||  
QY 424 GTTGATGTGTCGACTCAATAGCAATTAAGATGATGATGATGATGATGATGATGATGATGAT 365  
Db |||||  
QY 3446 TTTGTGGCAACACCTTAATCAGGAGCTTTCCAGATCTACAGTCTTATTAAGATGTATGC 3505  
Db |||||  
QY 364 TTTGTCGAGCCCGCAACAGGAGATGTTTCTGACATGCAATATTACTATGCAAGTGT 305  
Db |||||  
QY 3506 CTCCCTGTGTAATAGTACTATCTTAAACAAGTATGATGCTGTACCATGAGGTACGTGAT 3565  
Db |||||  
QY 304 TTGCGGGAACAGTACTATCTTAAATGAGTATGATGCTGTGAAACACCTCGAAACCT 245  
Db |||||  
QY 3566 AATAGTCTTAATGTAAGGATTTGTTCTTGTATTTTCCAAAGTATTTCCGATGCCAAAG 3625  
Db |||||  
QY 244 AATAATTTGAATGTCAAGGATTTGTTGATATGTCGAAGTCCGTGCTCTTCCGAGA 185  
Db |||||  
QY 3626 GAGGTGAACCATGCTAGAGCCAGTTTTCGTCACCGCGGACCCGCAAGGGCTGCA 3685  
Db |||||  
QY 184 GAATCTGAGACGACATTTGAACACCTGTGATCAGGACTGCTGTGAAACACCTCGAAACCT 125  
Db |||||  
QY 3686 GGACTACTCGAAATCTGTTGCAATGATTAAGAAATTTCAACGACACGACCTGACG 3745  
Db |||||  
QY 124 GGATGTTGGAATACTTTGGTGGCGATGATCAAAAGAAATTTCAACTCTCCGAAATTAATA 65  
Db |||||  
QY 3746 GGGAGATGATGACATGAGAGCAGCCGATCTGTTGTAGTAGATAGTGTATTTTGTAT 3799  
Db |||||  
QY 64 GGGGTGCTGTGACATCGAAGACACCGCTTCTCTAGTAGTAGATAGTATTTTGTAT 11  
Db |||||

Search completed: January 17, 2005, 16:28:44

Job time : 12565.7 secs